

Utah State University

DigitalCommons@USU

All Graduate Plan B and other Reports

Graduate Studies

1975

Multivariate Analysis of Variance for Simple Designs

Yin-Yin Chen

Utah State University

Follow this and additional works at: <https://digitalcommons.usu.edu/gradreports>



Part of the [Applied Statistics Commons](#)

Recommended Citation

Chen, Yin-Yin, "Multivariate Analysis of Variance for Simple Designs" (1975). *All Graduate Plan B and other Reports*. 1166.

<https://digitalcommons.usu.edu/gradreports/1166>

This Report is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Plan B and other Reports by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.



MULTIVARIATE ANALYSIS OF VARIANCE
FOR SIMPLE DESIGNS

by

Yin-Yin Chen

A report submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Applied Statistics

Plan B

Approved:

UTAH STATE UNIVERSITY
Logan, Utah

1975

TABLE OF CONTENTS

	Page
INTRODUCTION	1
OBJECTIVES	3
METHODOLOGY	4
NUMERICAL EXAMPLE	17
PROGRAM MODIFICATIONS	21
PROGRAM DOCUMENTATION	23
SAMPLE PROBLEMS	25
LITERATURE CITED	35
VITA	36

INTRODUCTION

The analysis of variance is a well known tool for testing how treatments change the average response of experimental units. The essence of the procedure is to compare the variation among means of groups of units subjected to the same treatment with the within treatment variation. If the variation among means is large with respect to the within group variation we are likely to conclude that the treatments caused the variation and hence we say the treatments cause some change in the group means.

The usual analysis of variance checks how far apart the group means are in a single scale of measurement. Almost all researchers are interested in how the treatments affect more than one characteristic (variable) of their experimental units. A typical usage of such data is to run a standard analysis of variance on each variable. This procedure can be very misleading when trying to interpret the results. Most of the time there are strong correlations among these variables and hence if one variable tests significant the others will also. The multivariate analysis of variance provides a way of performing valid tests regardless of the correlation structure among the variables of interest.

The multivariate analysis of variance compares the distance the treatments are apart in multidimensional space with the multidimensional

variance/covariance structure of the observations about the treatment means. Treatments that are far apart will likely be judged as being different.

Multivariate procedures are relatively unused for two primary reasons. The computational procedures are complex and there have been very few computer programs written to do them. This report is an attempt to encourage research workers to use more multivariate procedures.

OBJECTIVES

The report has two basic objectives:

1. To provide a computer program capable of performing the multivariate analysis of variance for the simple experimental designs; completely randomized, randomized block and Latin square designs.

2. To document the computer program and illustrate its usage.

The first objective will be met by adding multivariate capabilities to the program called BASIC that was developed by Greenhalgh (1967). The second objective will be met by making the necessary modification to the BASIC documentation and providing several illustrations of usage.

METHODOLOGY

Multivariate analysis of variance is a generalization of univariate analysis of variance. For the one-way univariate analysis of variance, the j th observation from i th population is assumed to be generated by the linear model $Y_{ij} = \mu + \tau_i + e_{ij}$ in which μ is an overall mean, τ_i is an effect due to i th treatment, and e_{ij} is a normal random variable with mean zero and variance σ^2 . All e_{ij} 's are assumed independently distributed. The hypothesis of equal population means $H_0: \mu_1 = \mu_2 \dots = \mu_g$ is an elementary case of the general linear hypothesis, because $\mu_i = \mu + \tau_i$ is the mean of i th treatment, so the hypothesis $H_0: \tau_1 = \tau_2 = \dots = \tau_g$ is equivalent. If we use matrix form to express all the observations then $Y = A\beta + \epsilon$, in which $Y' = [Y_{11}, \dots, Y_{1N_1}, \dots, Y_{g1}, \dots, Y_{gN_g}]$, $\epsilon = [e_{11}, \dots, e_{1N_1}, \dots, e_{g1}, \dots, e_{gN_g}]$ and parameter vector $\beta' = [\tau_1, \dots, \tau_g, \mu]$, where number of subjects in the i th group is N_i , and $N = N_1 + \dots + N_g$. A is the $N \times (g + 1)$ design matrix for the one-way analysis of variance model, the element of i th column and $g + 1$ column are 1, when i th group is applied. Postmultiplication of the design matrix by the parameter vector assures that ij th observation will involve only the constant $\mu + \tau_i$. The null hypothesis of the one-way analysis of variance can be expressed in matrix form as $H_0: C\beta = 0$.

Now extend the linear model and hypothesis to several dependent variates, i.e. multivariate. Let Y be p -dimensional multinormal random variable, and collect N observation vectors Y_1, Y_2, \dots, Y_N under some experimental design. The j th observation on the r th response is generated by this model $Y_{jr} = a_{ji} \xi_{1r} + \dots + a_{jm} \xi_{mr} + e_{jr}$, a_{jr} is the coefficient of each response in the j th vector, so the design matrix A is the same for all dimensions. e_{jr} , $r = 1, 2, \dots, p$ is the residual deviates of the j th observation. These are distributed with null mean vector and covariance matrix Σ of full rank p . The model for all observations using matrix form is $Y = A\xi + \epsilon$, Y is $N \times p$ matrix, has N observation vectors or rows, A is the appropriate design matrix. The matrix

$$\xi = \begin{bmatrix} \xi_{11} & \dots & \xi_{1p} \\ \xi_{21} & \dots & \xi_{2p} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \xi_{q1} & \dots & \xi_{qp} \end{bmatrix} \quad (q = g + 1)$$

is the unknown parameters matrix. ϵ is an $N \times p$ matrix and contains the residual variates e_{jr} .

The multivariate extension of the general linear hypothesis is

$H_0: C\xi M = 0$, in which C is a $(g-1) \times (g+1)$ matrix as in the univariate

case and refers to the hypothesis on the elements within given columns of the parameter matrix, M is $p \times r$ matrix and permits the generation of hypothesis among the different response parameters. The multivariate hypothesis is true if and only if the univariate hypothesis $H_0: C\xi Ma = 0$ holds for all nonnull r -component vectors a .

We will now use the general results to extend the analysis of variance, for some common experimental designs, to the multiple responses case.

(1) The one-way classification. (CRD)

There are g treatments (if the treatments were assigned at random then it is completely randomized designs), each treatment has p response variates on the sampling units. These measurements are assumed to be independent observations on p -dimensional multinormal variates with mean vectors $\mu_1, \mu_2, \dots, \mu_g$ under different treatments and a common unknown covariance Σ for all g conditions. The design matrix A in this model is the same as before, but its parameter matrix is

$$\xi = \begin{bmatrix} \tau_{11} & \tau_{12} & \cdots & \tau_{1p} \\ \tau_{21} & \tau_{22} & \cdots & \tau_{2p} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \tau_{g1} & \tau_{g2} & \cdots & \tau_{gp} \\ \mu_1 & \mu_2 & \cdots & \mu_p \end{bmatrix}$$

The hypothesis to be tested is that of equal-treatment-effect vectors:

$$H_0: \begin{bmatrix} \tau_{11} \\ \tau_{12} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \tau_{1p} \end{bmatrix} = \dots = \begin{bmatrix} \tau_{g1} \\ \tau_{g2} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \tau_{gp} \end{bmatrix}$$

which is the same as $H_0: C\xi M = 0$, where M is $p \times p$ identity matrix.

The matrix T of sums of squares and product among treatments, and the error matrix E will be found¹

$$t_{rs} = \sum_{i=1}^g \frac{1}{N_i} T_{ir} T_{is} - \frac{1}{N} G_r G_s$$

$$e_{rs} = \sum_{i=1}^g \sum_{j=1}^{N_i} Y_{ijr} Y_{ijs} - \sum_{i=1}^g \frac{1}{N_i} T_{ir} T_{is}$$

¹ It should be realized that the elements of these matrices can be calculated by the procedures given in elementary texts on the analysis of variance. Elements in the ii position, 11, 22, 33 etc. are the standard computations for the univariate analysis for the 1st, 2nd, 3rd response variates. The elements in the ij position, $i \neq j$, are computed in the same way as the cross product terms in the analysis of covariance. Sample problems given later will make these computations clear.

where Y_{ijr} = jth observation on response r under treatment i.

$T_{ir} = \sum_{j=1}^{N_i} Y_{ijr}$ = sum of all observations on rth response in presence of treatment i.

$G_r = \sum_{i=1}^g T_{ir}$ = grand total of all observations on rth response.

$$N = N_1 + N_2 + \dots + N_g$$

To test the equality mean vectors we must assume that the unknown variance matrix is common for all treatments. This assumption is another hypothesis we may test.

$$H_0: \Sigma_1 = \Sigma_2 = \dots = \Sigma_g$$

(2) Randomized blocks. (RBD)

In this model inferences from the observations will be restricted to the g treatments applied to just those b blocks employed in the experiment. The mathematical model for each observation is:

$$Y_{ijr} = \mu_r + \tau_{ir} + \beta_{jr} + e_{ijr}$$

where μ_r = usual general level effect for rth response.

τ_{ir} = effect of ith treatment on rth response.

β_{jr} = effect of jth block on rth response.

e_{ijr} = random effect specific to ijth combination of treatment, block and response.

The random terms $e'_{ij} = [e_{ij1}, e_{ij2}, \dots, e_{ijp}]$ are assumed to have the p -dimension multinormal distribution with null mean vector and common covariance matrix Σ for all combinations of i and j , and the e_{ij} in any block are independently distributed. The design matrix A is $(bg) \times (b + g + 1)$. The parameter matrix is

$$\xi = \begin{bmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1p} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \beta_{b1} & \beta_{b2} & \dots & \beta_{bp} \\ \tau_{11} & \tau_{12} & \dots & \tau_{1p} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \tau_{g1} & \tau_{g2} & \dots & \tau_{gp} \\ \mu_1 & \mu_2 & \dots & \mu_p \end{bmatrix}$$

The residual matrix ϵ has as rows the bg vectors e'_{ij} . In this model the usual hypothesis of interest is

$$H_0: \begin{bmatrix} \tau_{11} \\ \tau_{12} \\ \vdots \\ \tau_{1p} \end{bmatrix} = \dots = \begin{bmatrix} \tau_{g1} \\ \tau_{g2} \\ \vdots \\ \tau_{gp} \end{bmatrix}$$

of equal treatment effects. The T and E matrices of treatment and error can be found

$$t_{rs} = \frac{1}{b} \sum_{i=1}^g T_{ir} T_{is} - \frac{1}{bg} G_r G_s$$

$$e_{rs} = \sum_{i=1}^g \sum_{j=1}^b Y_{ijr} Y_{ijs} - \frac{1}{b} \sum_{i=1}^g T_{ir} T_{is} - \frac{1}{g} \sum_{j=1}^b B_{jr} B_{js} + \frac{1}{bg} G_r G_s$$

where $B_{jr} = \sum_{i=1}^g Y_{ijr}$ = the total of the observations on the rth response in block j.

$$T_{ir} = \sum_{j=1}^b Y_{ijr} = \text{the total of the rth observations under treatment } i.$$

$$G_r = B_{1r} + B_{2r} + \dots + B_{br} = T_{1r} + T_{2r} + \dots + T_{gr}$$

= grand total of the values of that response in all sampling units.

(3) Latin square. (LSD)

The idea of a square is evident, if g treatments are to be investigated, it has g^2 experimental units. The mathematical model for each observation is

$$Y_{ijk} = \mu_r + \alpha_{jr} + \beta_{kr} + \tau_{ir} + e_{ijk}$$

where μ_r = general level parameter of r th response.

α_{jr} = effect of j th row treatment on r th response.

β_{kr} = effect of k th column treatment on r th response.

τ_{ir} = effect of i th treatment on r th response.

e_{ijk_r} = usual multinormal random variable term.

The design matrix A and parameter matrix ξ can be produced by using the same methods as for randomized block design. The null hypothesis

$$H_0: \begin{bmatrix} \tau_{11} \\ \tau_{12} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \tau_{1p} \end{bmatrix} = \dots = \begin{bmatrix} \tau_{g1} \\ \tau_{g2} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \tau_{gp} \end{bmatrix}$$

is the usual hypothesis to be tested in Latin square design. It is not possible to test any hypothesis concerning rows or columns². Proceeding to the test of H_0 , we calculate matrices T and E for treatment and error.

²This is for the same reason we do not test the effects of blocks in the randomized block design. These design effects are created by the researcher by the manner in which he restricts his randomization, selects his experimental units. A test, if made, has no probability associated with it. Large values of the mean squares simply mean that the researcher did a good job in design.

$$t_{rs} = \frac{1}{g} \sum_{i=1}^g T_{ir} T_{is} - \frac{1}{g^2} G_r G_s$$

$$e_{rs} = \sum_{i=1}^g \sum_{j=1}^g \sum_{k=1}^g Y_{ijkr} Y_{ijks} - \frac{1}{g} \sum_{j=1}^g R_{jr} R_{js} - \frac{1}{g} \sum_{k=1}^g C_{kr} C_{ks} -$$

$$\frac{1}{g} \sum_{i=1}^g T_{ir} T_{is} + \frac{2}{g^2} G_r G_s$$

where

$$R_{jr} = \sum_{i=1}^g \sum_{k=1}^g Y_{ijkr}$$

$$C_{kr} = \sum_{i=1}^g \sum_{j=1}^g Y_{ijkr}$$

$$T_{ir} = \sum_{j=1}^g \sum_{k=1}^g Y_{ijkr}$$

$$G_r = R_{1r} + R_{2r} + \dots + R_{gr} = C_{1r} + C_{2r} + \dots + C_{gr} =$$

$$T_{1r} + T_{2r} + \dots + T_{gr}$$

The test for equality of dispersion, $H_0: \Sigma_1 = \Sigma_2 = \dots = \Sigma_g$, can only be tested on the completely randomized design. The test for equality of mean vectors, $H_0: \mu_1 = \mu_2 = \dots = \mu_g$, can be tested on all designs. Here are the procedures for these two possible tests:

(i) Hypothesis of equality of the mean vectors.

$$H_0: \mu_1 = \mu_2 = \dots = \mu_g$$

Using Rao's (1952) notation

$$\Lambda = \frac{|E|}{|T + E|}$$

where E is the within-treatment deviation SSSP matrix of error, T is the SSSP matrix of treatment. In completely randomized design, the T + E matrix equals the total variance/covariance matrix.

The notations that will be used are the following:

g = number of groups or treatments.

N_i = number of subjects in group (i).

N = total number of subjects.

q = g - 1 = degree of freedom for treatment.

p = number of responses.

t = $DF_{\text{treatment}} + DF_{\text{error}} - k$, where DF means degree of freedom, k = number of covariates.

Let

$$s = \sqrt{\frac{\frac{p^2 q^2 - 4}{2^2}}{p + q - 5}}$$

$$m = t - \frac{p + q + 1}{2}$$

and $\lambda = \frac{pq - 2}{4}$

then

$$(1) F = \frac{1-\Lambda}{\Lambda} \frac{t-p}{p} \text{ with DF } (p, t-p); \text{ if } q = 1, \text{ for any } p$$

$$(2) F = \frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}} \frac{t-p-1}{p} \text{ with DF } (2p, 2(t-p-1)); \text{ if } q=2, \text{ for any } p$$

$$(3) F = \frac{1-\Lambda}{\Lambda} \frac{t-q}{q} \text{ with DF } (q, t-q); \text{ if } p = 1, \text{ for any } q$$

$$(4) F = \frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}} \frac{t-q-1}{q} \text{ with DF } (2q, 2(t-q-1)); \text{ if } p=2, \text{ for any } q$$

$$(5) F = \frac{1-\Lambda^{\frac{1}{s}}}{\Lambda^{\frac{1}{s}}} \frac{ms-2\lambda}{pq} \text{ with DF } (pq, (ms-2\lambda)); p, q > 2$$

Approximate F not integer DF

(ii) Hypothesis of the equality of group dispersions (CRD only).

$$H_0: \Sigma_1 = \Sigma_2 = \dots = \Sigma_g$$

This is presented by Box (1949), Box defines the criterion M:

$$\begin{aligned} M &= (N-g) \ln \left| \frac{E}{N-g} \right| - \sum_i (N_i-1) \ln \left| \frac{E_i}{N_i-1} \right| \\ &= (N-g) \ln |S| - \sum_i (N_i-1) \ln |S_i| \end{aligned}$$

where S = the pooled matrix among treatments.

S_i = variance/covariance matrix within treatment i.

Required constants are:

$$A_1 = \left(\sum_i \frac{1}{N_i - 1} - \frac{1}{N - g} \right) \frac{2p^2 + 3p - 1}{6(p+1)(g-1)}$$

$$A_2 = \left(\sum_i \frac{1}{(N_i - 1)^2} - \frac{1}{(N - g)^2} \right) \frac{(p-1)(p+2)}{6(g-1)}$$

if $A_2 - A_1^2$ is positive, then

$$n_1 = \frac{(g-1)p(p+1)}{2} \quad n_2 = \frac{n_1 + 2}{A_2 - A_1^2}$$

$$b = \frac{n_1}{1 - A_1 - n_1/n_2}$$

$$F(n_1, n_2) = \frac{M}{b}$$

if $A_2 - A_1^2$ is negative, then

$$n_1 = \frac{(g-1)p(p+1)}{2} \quad n_2 = \frac{n_1 + 2}{A_1^2 - A_2}$$

$$b = \frac{n_2}{1 - A_1 + 2/n_2}$$

$$F(n_1, n_2) = \frac{n_2 M}{n_1(b - M)}$$

NUMERICAL EXAMPLE

This is a randomized block design with three treatments, four blocks, and two variables. Use this data set to do the multivariate F test.

treatment	block			
	1	2	3	4
1	4 8	5 7	8 8	5 11
2	6 6	1 8	2 7	2 6
3	3 2	5 12	9 11	3 13

Use the notation Y_{ijk} = observation in i th treatment, j th block, k th response.

$$\sum_j \sum_i Y_{ij1}^2 = 299$$

$$\sum_j \sum_i Y_{ij2}^2 = 921$$

$$\sum_j \sum_i Y_{ij1} Y_{ij2} = 460$$

$$\sum_j Y_{1j1} = 22$$

$$\sum_j Y_{2j1} = 11$$

$$\sum_j Y_{3j1} = 20$$

$$\sum_i \sum_j Y_{ij1} = 53$$

$$\sum_j Y_{1j2} = 34$$

$$\sum_j Y_{2j2} = 27$$

$$\sum_j Y_{3j2} = 38$$

$$\sum_i \sum_j Y_{ij2} = 99$$

$$\sum_i Y_{i11} = 13$$

$$\sum_i Y_{i21} = 11$$

$$\sum_i Y_{i31} = 19$$

$$\sum_i Y_{i41} = 10$$

$$\sum_i Y_{i12} = 16$$

$$\sum_i Y_{i22} = 27$$

$$\sum_i Y_{i32} = 26$$

$$\sum_i Y_{i42} = 30$$

$$\text{Corrected total for } Y_1^2 = \sum \sum Y_{ij1}^2 - \frac{(\sum \sum Y_{ij1})^2}{bg} = 299 - \frac{53^2}{12} = 299 - 234.083$$

$$DF = 11 \qquad = 64.917$$

$$Y_2^2 = \sum \sum Y_{ij2}^2 - \frac{(\sum \sum Y_{ij2})^2}{bg} = 921 - \frac{99^2}{12} = 921 - 816.75$$

$$= 104.25$$

$$Y_1 Y_2 = \sum \sum Y_{ij1} Y_{ij2} - \frac{(\sum \sum Y_{ij1})(\sum \sum Y_{ij2})}{bg} = 460 - \frac{53 \times 99}{12}$$

$$= 460 - 437.25 = 22.75$$

$$\text{Corrected treatments for } Y_1^2 = \frac{\sum (\sum Y_{ij1})^2}{b} - \frac{(\sum \sum Y_{ij1})^2}{bg} = \frac{22^2 + 11^2 + 20^2}{4} - \frac{53^2}{12}$$

$$DF = 2 \qquad = 251.25 - 234.083 = 17.167$$

$$Y_2^2 = \frac{\sum (\sum Y_{ij2})^2}{b} - \frac{(\sum \sum Y_{ij2})^2}{bg} = \frac{34^2 + 27^2 + 38^2}{4} - \frac{99^2}{12}$$

$$= 832.25 - 816.75 = 15.5$$

$$Y_1 Y_2 = \frac{\sum (\sum Y_{ij1})(\sum Y_{ij2})}{b} - \frac{(\sum \sum Y_{ij1})(\sum \sum Y_{ij2})}{bg} = \frac{22 \times 34 + 11 \times 27 + 20 \times 38}{4}$$

$$- \frac{53 \times 99}{12}$$

$$= 451.25 - 437.25 = 14.0$$

$$\text{Corrected blocks for } Y_1^2 = \frac{\sum (\sum Y_{ij1})^2}{g} - \frac{(\sum \sum Y_{ij1})^2}{bg} = \frac{13^2 + 11^2 + 19^2 + 10^2}{3} - \frac{53^2}{12}$$

$$DF = 3 \qquad = 250.333 - 234.083 = 16.25$$

$$Y_2^2 = \frac{\Sigma(\Sigma Y_{ij2})^2}{g} - \frac{(\Sigma \Sigma Y_{ij2})^2}{bg} = \frac{16^2 + 27^2 + 26^2 + 30^2}{3} - \frac{99^2}{12}$$

$$= 853.667 - 816.75 = 36.917$$

$$Y_1 Y_2 = \frac{\Sigma(\Sigma Y_{ij1})(\Sigma Y_{ij2})}{g} - \frac{(\Sigma \Sigma Y_{ij1})(\Sigma \Sigma Y_{ij2})}{bg}$$

$$= \frac{16 \times 13 + 11 \times 27 + 19 \times 26 + 10 \times 30}{3} - \frac{53 \times 99}{12}$$

$$= 433 - 437.25 = -4.25$$

In matrix form:

$$\text{Total: } \begin{bmatrix} 64.917 & 22.75 \\ & 104.25 \end{bmatrix} \quad \text{Treatments: } \begin{bmatrix} 17.167 & 14.0 \\ & 15.5 \end{bmatrix}$$

$$\text{Blocks: } \begin{bmatrix} 16.25 & -4.25 \\ & 36.917 \end{bmatrix}$$

Error = Total - Treatments - Blocks:

$$DF = 6 \begin{bmatrix} 64.917 & 22.75 \\ & 104.25 \end{bmatrix} - \begin{bmatrix} 17.167 & 14.0 \\ & 15.5 \end{bmatrix} - \begin{bmatrix} 16.25 & -4.25 \\ & 36.917 \end{bmatrix} = \begin{bmatrix} 31.5 & 13.0 \\ & 51.833 \end{bmatrix}$$

$$T + E = \begin{bmatrix} 17.167 & 14.0 \\ & 15.5 \end{bmatrix} + \begin{bmatrix} 31.5 & 13.0 \\ & 51.833 \end{bmatrix} = \begin{bmatrix} 48.667 & 27.0 \\ & 67.333 \end{bmatrix}$$

$$\Lambda = \frac{|E|}{|T + E|} = \frac{1463.75}{2547.89} = 0.5745$$

$$p = 2. \quad q = 2. \quad t = 2 + 6 = 8$$

$$\sqrt{\Lambda} = 0.758$$

$$F = \frac{1 - \sqrt{\Lambda}}{\sqrt{\Lambda}} \frac{t - p - 1}{p} = \frac{1 - 0.758}{0.758} \times \frac{8 - 2 - 1}{2} = \frac{0.242}{0.758} \times 2.5$$

$$= 0.32 \times 2.5 = 0.80$$

associated with $DF = (2p, 2(t-p-1))$ ie. (4, 10)

The computer output for the same data is given in Sample Problem 1.

PROGRAM MODIFICATIONS

The STATPAC/BASIC program was written using matrix mode. In performing the desired computations on multiple variables all the matrices needed for the multivariate tests were produced but only the main diagonals used. To modify the program the matrices were captured on disk and held for later use. The only thing we have to note is it holds all the variance/covariance matrices within each treatment on disk as an extra in the completely randomized design.

Following basic analysis of variances the program will perform the multivariate tests. If the model is completely randomized design, read all variance/covariance matrices within treatment (S_i) from disk, then calculate the pooled matrix among treatments (S). These matrices are used for testing the hypothesis of equality of treatment dispersion $H_0: \Sigma_1 = \Sigma_2 = \dots = \Sigma_g$. The program branches to FUNCTION DDET to get the determinant of all S_i and S . Following the steps Box has presented in 1949, calculate M , A_1 , A_2 , n_1 , n_2 , b and F ratio for this test. If the model is randomized block design or Latin square design, we don't need to test this hypothesis, so skip this part in program.

The next step is to do the multivariate F test for the equality of treatment mean vectors. $H_0: \mu_1 = \mu_2 = \dots = \mu_g$ for all the models. Read all the SSSP matrices for the desired model from disk. Move the

error SSSP matrix to lower triangular portion of matrix A and move the sum of error and treatment SSSP matrix to upper triangular portion of matrix A. Branch to SUBROUTINE MULTF. SUBROUTINE MULTF is the subprogram for testing the equality of the mean vectors which use Rao's notation. For different conditions of treatment and variable numbers, do the different F tests.

PROGRAM DOCUMENTATION

Up until now the BASIC program could calculate the analysis of variance or covariance for completely randomized design (CRD) with unequal sample size, or randomized block design (RBD), or Latin square design (LSD) without replications or subsampling. With covariance, on a CRD, it will, on control, give a linear regression analysis within each treatment. This program can now do the multivariate procedures with or without covariance on CRD, RBD and LSD.

Automatically the program gives an analysis of variance and treatment means for each variable. With covariance, it gives the error correlation matrix, inverse matrix, solution matrix, adjusted analysis of variance, and adjusted means. Following these with multivariate control are variance/covariance matrix within each treatment, pooled matrix for treatment (for CRD only). The SSSP matrix of row, column for LSD; block for RBD; and SSSP matrix of treatment and error for all of the designs.

The control card for STATPAC/BASIC program is now:

<u>Column</u>	<u>Description</u>
(4)	Model identification, 1 = Completely randomized design 2 = Randomized block design

SAMPLE PROBLEMS

Two sample problems will be used to demonstrate the capabilities of this program. The first is randomized block design with three treatments, four blocks. It has two responses per experimental unit (the same data as numerical example). The control card for this problem is:

<u>Column</u>	<u>Description</u>
(4)	2 = RBD
(8)	3 = number of treatments
(12)	4 = number of block
(16)	0 = no covariate
(20)	2 = number of Y's
(22)	0 = RBD can't do regression within treatment
(24)	5 = card reader
(26)	1 = output multivariate test
(41 - 80)	descriptive information

The second sample is a randomized block design with twenty-four treatments, two blocks. It has four responses per experimental unit.

The control card is as follows:

<u>Column</u>	<u>Description</u>
(4)	2 = RBD
(7-8)	24 = number of treatments
(12)	2 = number of blocks
(16)	0 = no covariate
(20)	4 = number of Y's
(22)	0 = RBD can't do regression within treatment
(24)	5 = card reader
(26)	1 = output multivariate test
(41-80)	descriptive information

Sample problem 1 (input cards)

1	2	3	4	2	5	1
2	(4X,2F2.0)					
3	1	2				
4	1	1	4	8		
5	1	2	5	7		
6	1	3	8	8		
7	1	4	5	1	1	
8	2	1	6	6		
9	2	2	1	8		
10	2	3	2	7		
11	2	4	2	6		
12	3	1	3	2		
13	3	2	5	1	2	
14	3	3	9	1	1	
15	3	4	3	1	3	

EXAMPLE

Sample problem 1

Computer output of hand calculated example

2 3 4 0 2 0 5 1
 (4X,2F2.0)
 1 2

EXAMPLE

ANALYSIS OF VARIANCE, VARIABLE 1

SOURCE	DF	SS	MS	F
TOT	11	.6491567E+02		
BLK	3	.1625000E+02		
TRT	2	.1716667E+02	.8583333E+01	.1634921F+01
FRR	6	.3150000E+02	.5250000E+01	

TRT	TRT MEANS	SE
1	.5500000E+01	.1145644F+01
2	.2750000E+01	.1145644F+01
3	.5000000E+01	.1145644F+01

EXP MEAN .44166667E+01 C.V. .51878215E+00

ANALYSIS OF VARIANCE, VARIABLE 2

SOURCE	DF	SS	MS	F
TOT	11	.1042500E+03		
BLK	3	.3691567E+02		
TRT	2	.1550000E+02	.7750000E+01	.8971061F+00
FRR	6	.5183333E+02	.8636889E+01	

TRT	TRT MEANS	SE
1	.8500000E+01	.1469599F+01
2	.6750000E+01	.1469599F+01
3	.9500000E+01	.1469599F+01

EXP MEAN .82500000E+01 C.V. .35626651E+00

BLOCK SSSP MATRIX WITH 3. DF
 1 .1625000E+02 -.4250000E+01
 2 .3691667E+02

TREATMENT SSSP MATRIX WITH 2. DF
 1 .1716667E+02 .1400000E+02
 2 .1550000E+02

ERROR SSSP MATRIX WITH 6. DF
1 .3150000E+02 .1300000E+02
2 .5183333E+02

TREAT. + ERROR SSSP MATRIX IS
1 .4866667E+02 .2700000E+02
2 .6733333E+02

T+E DET= .2547889E+04
E DFT= .1463750E+04

MULTIVARIATE F= 0.80 WITH 4.0 10.0 DF

Sample problem 2 (input cards)

2 24 2 0 4 0 5 1
 (6X,4F3.0)
 1 2 3 4

RBD WITHOUT COVARIANCE.

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

Sample problem 2 (computer output)

2 24 2 0 4 0 5 1
 (6X,4F3.0)
 1 2 3 4

RBD WITHOUT COVARIANCE

ANALYSIS OF VARIANCE, VARIABLE 1

SOURCE	DF	SS	MS	F
TOT	47	.5196479E+04		
BLK	1	.5852083E+02		
TRT	23	.3767974E+04	.1636252E+03	.2750392E+01
FRR	23	.1369979E+04	.5956431E+02	

TRT	TRT MEANS	SE
1	.2000000E+01	.5457303E+01
2	.8000000E+01	.5457303E+01
3	.1150000E+02	.5457303E+01
4	.4500000E+01	.5457303E+01
5	.3000000E+01	.5457303E+01
6	.6500000E+01	.5457303E+01
7	.3500000E+01	.5457303E+01
8	.5000000E+01	.5457303E+01
9	.1000000E+01	.5457303E+01
10	.9500000E+01	.5457303E+01
11	.9000000E+01	.5457303E+01
12	.1000000E+02	.5457303E+01
13	.6500000E+01	.5457303E+01
14	.8000000E+01	.5457303E+01
15	.7500000E+01	.5457303E+01
16	.7000000E+01	.5457303E+01
17	.9000000E+01	.5457303E+01
18	.1150000E+02	.5457303E+01
19	.2200000E+02	.5457303E+01
20	.2600000E+02	.5457303E+01
21	.2800000E+02	.5457303E+01
22	.3500000E+02	.5457303E+01
23	.1050000E+02	.5457303E+01
24	.2500000E+02	.5457303E+01

EXP MEAN .11229167E+02 C.V. .68729872E+00

ANALYSIS OF VARIANCE, VARIABLE 2

SOURCE	DF	SS	MS	F
TOT	47	.7396667E+03		
BLK	1	.2133333E+02		
TRT	23	.4016667E+03	.1746377E+02	.1268421E+01
FRR	23	.3166667E+03	.1376812E+02	

TRT	TRT MEANS	SE
1	.5000000E+01	.2623749F+01
2	.1050000E+02	.2623749F+01
3	.1000000E+02	.2623749F+01
4	.8500000E+01	.2623749F+01
5	.1050000E+02	.2623749F+01
6	.9500000E+01	.2623749F+01
7	.4500000E+01	.2623749F+01
8	.7000000E+01	.2623749F+01
9	.7000000E+01	.2623749F+01
10	.6000000E+01	.2623749F+01
11	.6500000E+01	.2623749F+01
12	.9000000E+01	.2623749F+01
13	.7000000E+01	.2623749F+01
14	.9000000E+01	.2623749F+01
15	.7000000E+01	.2623749F+01
16	.8500000E+01	.2623749F+01
17	.8500000E+01	.2623749F+01
18	.8500000E+01	.2623749F+01
19	.5000000E+01	.2623749F+01
20	.5000000E+00	.2623749F+01
21	.5000000E+00	.2623749F+01
22	.7000000E+01	.2623749F+01
23	.1150000E+02	.2623749F+01
24	.6000000E+01	.2623749F+01

EXP MEAN .69166667E+01 C.V. .53646379E+00

ANALYSIS OF VARIANCE, VARIABLE 3

SOURCE	DF	SS	MS	F
TCT	47	.8826667E+03		
BLK	1	.7500000E+00		
TRT	23	.3336667E+03	.1450725E+02	.6086031E+00
ERR	23	.5482500E+03	.2383696E+02	

TRT	TRT MEANS	SE
1	.7500000E+01	.3452315F+01
2	.8000000E+01	.3452315F+01
3	.6500000E+01	.3452315F+01
4	.1050000E+02	.3452315F+01
5	.4000000E+01	.3452315F+01
6	.5500000E+01	.3452315F+01
7	.8500000E+01	.3452315F+01
8	.9000000E+01	.3452315F+01
9	.6500000E+01	.3452315F+01
10	.8500000E+01	.3452315F+01
11	.7000000E+01	.3452315F+01
12	.9500000E+01	.3452315F+01

13	.6500000E+01	.3452315F+01
14	.1050000E+02	.3452315F+01
15	.1150000E+02	.3452315F+01
16	.1000000E+02	.3452315F+01
17	.1400000E+02	.3452315F+01
18	.1100000E+02	.3452315F+01
19	.1050000E+02	.3452315F+01
20	.1050000E+02	.3452315F+01
21	.1000000E+02	.3452315F+01
22	.2000000E+01	.3452315F+01
23	.7500000E+01	.3452315F+01
24	.5000000E+01	.3452315F+01

EXP MEAN .8333333F+01 C.V. .58587727E+00

ANALYSIS OF VARIANCE, VARIABLE 4

SOURCE	DF	SS	MS	F
TOT	47	.3729979E+04		
BLK	1	.1725208E+03		
TRT	23	.2532479E+04	.1101078E+03	.2470762E+01
FRR	23	.1024979E+04	.4456431E+02	

TRT	TRT MEANS	SE
1	.2550000E+02	.4720398F+01
2	.1350000E+02	.4720398F+01
3	.1200000E+02	.4720398F+01
4	.1650000E+02	.4720398F+01
5	.2250000E+02	.4720398F+01
6	.1850000E+02	.4720398F+01
7	.2350000E+02	.4720398F+01
8	.1900000E+02	.4720398F+01
9	.2550000E+02	.4720398F+01
10	.1600000E+02	.4720398F+01
11	.1750000E+02	.4720398F+01
12	.1150000E+02	.4720398F+01
13	.2000000E+02	.4720398F+01
14	.1250000E+02	.4720398F+01
15	.1400000E+02	.4720398F+01
16	.1450000E+02	.4720398F+01
17	.1050000E+02	.4720398F+01
18	.9000000E+01	.4720398F+01
19	.2500000E+01	.4720398F+01
20	.3000000E+01	.4720398F+01
21	.1500000E+01	.4720398F+01
22	.1000000E+01	.4720398F+01
23	.1050000E+02	.4720398F+01
24	.4000000E+01	.4720398F+01

EXP MEAN .1352083F+02 C.V. .49373071E+00

BLOCK SSSP MATRIX WITH 1. DF

1	.5452083E+02	.3533333E+02	.6625000E+01	-.1004792E+03
2	.2133333E+02	.4000000E+01	-.6066667E+02	
3	.7500000E+00	-.1137500E+02		
4	.1725208E+03			

TREATMENT SSSP MATRIX WITH 23. DF

1	.3767979E+04	-.7935833E+03	-.1686667E+03	-.2805729E+04
2	.4016667E+03	-.2316667E+02	.4150833E+03	
3	.3336667E+03	-.1418333E+03		
4	.2532479E+04			

ERROR SSSP MATRIX WITH 23. DF

1	.1369979E+04	-.9483333E+02	-.4946250E+03	-.7805208E+03
2	.3166667E+03	-.1550000E+02	-.2063333E+03	
3	.5482500E+03	-.3812500E+02		
4	.1024979E+04			

TREAT. + ERROR SSSP MATRIX IS

1	.5137958E+04	-.8884167E+03	-.6632917E+03	-.3586250E+04
2	.7183333E+03	-.3866667E+02	.2087500E+03	
3	.8819167E+03	-.1799583E+03		
4	.3557458E+04			

T+E DET= .1256464E+02
E DET= .8180896E-01

MULTIVARIATE F= 2.28 WITH 92.0 81.7 DF

LITERATURE CITED

- Box, G. E. P. 1949. A General Distribution Theory for a Class of Likelihood Criteria, *Biometrika*, Vol. 36, pp. 317-346.
- Cooley, William W., and Paul R. Lohnes. 1962. *Multivariate Procedures for the Behavioral Sciences*. John Wiley and Sons, Inc., New York. 207 p.
- Greenhalgh, H. Wain. 1967. FORTRAN Programs for the Calculation of Most of the Commonly used Experimental design Models. Unpublished M.S. Thesis. Utah State University.
- Morrison, Donald F. 1967. *Multivariate Statistical Methods*. McGraw-Hill Book Company, New York. 327 p.
- Ostle, Bernard. 1954. *Statistics in Research*. Iowa State University Press, Ames, Iowa. 480 p.
- Rao, C. R. 1952. *Advanced Statistical Methods in Biometric Research*. John Wiley and Sons, Inc., New York.
- Smith, H., R. Gnanadesikan, and J. B. Hughes. 1962. Multivariate analysis of variance (MANOVA), *Biometrics*, Vol. 18, pp.22-41.

VITA

Yin-Yin Chen

Candidate for the Degree of

Master of Science

Report: Multivariate Analysis of Variance for Simple Designs

Major Field: Applied Statistics

Biographical Information:

Personal Data: Born at Mainland, Republic of China, April 20, 1949, daughter of Chun-Ruey and Chun-Ying Chen.

Education: Received the Bachelor of Science degree from Tamkang College of Arts and Sciences with a major in Mathematics in June 1971; completed requirements for the Master of Science degree at Utah State University in 1975.

Professional Experience: From September 1971 to August 1974, Teaching Assistant in the Department of Computer Science, Tamkang College of Arts and Sciences, Taiper, Taiwan, Republic of China.