AD-A955 424

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a.	1b. RESTRICTIVE MARKINGS DTIC FILE COPY
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited. (2)
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 89-0488
6a. NAME OF PERFORMING ORGANIZATION Texas A&M University Institute of Statistics 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AFOSR
6c. ADDRESS (City, State, and ZIP Code) College Stations, Texas 77843	7b. ADDRESS (City, State, and ZIP Code) BLDG 410 BAFB DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR 74-2676
8c. ADDRESS (City, State, and ZIP Code) BLDG 410 BAFB DC 20332-6448	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 61102F PROJECT NO. 9769 TASK NO. 05 WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
RELIABILITY AND OTHER AIR FORCE PROBLEMS INVOLVING FACTORIZATIONS OF MULTIVARIATE TEST CRITERIA

12. PERSONAL AUTHOR(S)
A.M. Kshiragar

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM ______ TO ______	May 1976	39

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

DTIC
ELECTE
APR 26 1989
S D
D

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr. Charles Holland	22b. TELEPHONE (Include Area Code) 767-5025 22c. OFFICE SYMBOL NM

DD Form 1473, JUN 86 Previous editions are obsolete. SECURITY CLASSIFICATION OF THIS PAGE

89 4 25 008

AFOSR-TR- 89-0488

RELIABILITY AND OTHER AIR FORCE PROBLEMS INVOLVING FACTORIZATIONS OF MULTIVARIATE TEST CRITERIA

by

A. M. Kshirsagar

Institute of Statistics

Texas A&M University

Final Report

AF OSR-74-2676

Project 9769-05

Texas A&M Research Foundation

Project No. 3065

May 1976

ABSTRACT

This report is concerned with derivations of Hotelling's T_o^2, Pillai's V and Roy's maximum root type statistics corresponding to the direction and dimensionality statistics based on Wilks' Λ for testing the adequacy of s hypothetical discriminant functions. The proposed functions may either be in the p-dimensional $\underline{x}$-space or the q-dimensional $\underline{y}$-space. The distributions of these test statistics are obtained when $q < p$ and the functions are given in the $\underline{x}$-space, and also when $(q - s) < p$ and the functions are given in the dummy variables' space. The expressions for these statistics in terms of the cell frequencies of a contingency table are also given, for investigating the nature of association between two attributes.

Final Report of Project No. 3065

Contents

1. INTRODUCTION AND SUMMARY OF THE RESEARCH WORK COMPLETED UNDER THE PROJECT

Description of the Problem

Consider k random samples of sizes $n_1, n_2, \ldots, n_k$ from $k = q + 1$ independent p-variate normal populations $\pi_i (i = 1, 2, \ldots, k)$ with means $\underline{\mu}_i$ and the same variance-covariance matrix Σ. If $\underline{x}' = [x_1, x_2, \ldots, x_p]$ denotes the vector of p variables on which measurements are made, then to analyze the differences among groups or populations one forms the following multivariate analysis of variance table:

TABLE 1

MANOVA

Source	d.f.	S.S. & S.P. Matrix
Between groups	q	B_x
Within groups	$n - q$	W_x
Total	$n = n_1 + n_2 + \ldots + n_k - 1$	$W_x + B_x$

If one defines a q-variate vector $\underline{y}' = [y_1, y_2, \ldots, y_q]$ where $y_i (i = 1, 2, \ldots, q)$ are pseudo or dummy variables taking the values of one when an observation comes from the i^{th} group and zero otherwise, then B_x can be shown (see Kshirsagar [1972]) to be the matrix of regression sum of squares and sum of products (s.s. and s.p.) of $\underline{x}$ on $\underline{y}$. In other words,

$$B_x = C_{xy} C_{yy}^{-1} C_{yx} ,$$

where C_{xy} is the matrix of corrected sum of products of $\underline{x}$ and $\underline{y}$ and C_{yy} is the matrix of corrected s.s. and s.p. of $\underline{y}$. Furthermore, if C_{xx} is the matrix of the corrected s.s. and s.p. of $\underline{x}$, then

$$W_x = C_{xx.y} = C_{xx} - C_{xy} C_{yy}^{-1} C_{yx} .$$

The problem of discrimination among the groups then reduces to the investigation of the relationship between a p-variate vector $\underline{x}$ and a q-variate vector $\underline{y}$. The adequate number of discriminant functions required to describe this relationship is just the dimensionality of the k means $\underline{\mu}_i$. This is also the rank of the non-centrality matrix

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

where

$$\bar{\underline{\mu}} = \frac{1}{N} \sum_{i=1}^{k} n_i \underline{\mu}_i$$

and

$$N = \sum_{i=1}^{k} n_i .$$

If $r_1 > r_2 > \dots > r_f$ ($f = \min(p, q)$) are roots of the equation

$$|-r_i^2(W_x + B_x) + B_x| = 0 ,$$

then the rank of Ω is estimated by the number of significant roots, s, and $\underline{\ell}_i'\underline{x}$ ($i = 1, 2, \dots, s$), the corresponding canonical variables form a set of discriminant functions that adequately describe the relationship between $\underline{x}$ and $\underline{y}$. The vectors $\underline{\ell}_i$ satisfy

$$[-r_i^2(W_x + B_x) + B_x]\ \underline{\ell}_i = \underline{0} , \qquad i = 1, 2, \dots, s .$$

The remaining canonical variables $\underline{\ell}_{s+1}'\underline{x}$, $\underline{\ell}_{s+2}'\underline{x}$, ..., $\underline{\ell}_f'\underline{x}$ corresponding to the insignificant roots are called "null functions". Canonical variables of the $\underline{y}$-space corresponding to r_1^2, r_2^2, ..., r_s^2 are called discriminant functions from the space of the dummy variables and they define contrasts among $\underline{\mu}_1$, $\underline{\mu}_2$, ..., $\underline{\mu}_k$ that are significant. If $s = 1$, the means $\underline{\mu}_i$ of the k groups are collinear and only one discriminant function $\underline{\ell}_1'\underline{x}$, corresponding to r_1^2, is necessary to describe the relationship. If $s = 2$, then the group means are coplanar and $\underline{\ell}_1'\underline{x}$ and $\underline{\ell}_2'\underline{x}$ are adequate, etc.

Sometimes it is of interest to test the adequacy of one or more hypothetical discriminant functions. This has been discussed by Williams [1952a], Bartlett [1951] and others. The hypothesis, H, of goodness of fit of one or more hypothetical discriminant functions can be split into two parts.

H: (i) Collinearity or more generally dimensionality aspect or whether the number of assigned functions is adequate at all and

(ii) direction aspect or whether the functions that really discriminate adequately agree with the hypothesized ones.

So far it seems that the only tests constructed for testing the hypothesis, H, are based on Wilks' Λ criterion [1932]. Wilks' Λ has been factored into three parts (i) testing the discriminating ability of the hypothetical function(s), (ii) the collinearity factor and (iii) the direction factor. Furthermore, the three tests are independent under the null hypothesis. In this report we shall be concerned with finding the corresponding tests of adequacy of one or more assigned discriminant functions for the three other multivariate test criteria, viz.

(i) Hotelling's [1951] generalized T_o^2, $T_o^2 = \mathrm{tr}\, B_x W_x^{-1}$,

(ii) Pillai's [1955] V criterion, $V = \mathrm{tr}\, B_x(W_x + B_x)^{-1}$, and

(iii) Roy's [1957] maximum root criterion r_1^2, the largest root of $|-r^2(W_x + B_x) + B_x| = 0$.

The main reason for investigating these other criteria is that at present no adequate theory exists to guide the choice of a statistic from among Wilks' Λ, Hotelling's T_o^2, Pillai's V, and Roy's r_1^2.

Comparisons of all these criteria, from the point of view of their power against alternative hypotheses, have been done by Mikhail [1965], Pillai and Jayachandran [1967; 1968], Pillai and Dotson [1969] and Ito [1960]. It seems that the criteria are almost equivalent to each other and perform equally well, but these results have been based on Monte Carlo studies. In particular cases, one or the other of the criteria may have an edge over the others.

In this report, we consider the factors of Wilks' Λ as given by Kshirsagar [1970] for hypothetical functions from the $\underline{x}$-space. Since he expressed these factors as $|P|/|P + Q|$, where P and Q are independent Wishart matrices, the corresponding Hotelling's generalized T_o^2 and Pillai's V are easily seen to be $\text{tr } QP^{-1}$ and $\text{tr } Q(P + Q)^{-1}$, respectively. The only problem arises in that P and Q contain unknown quantities. Therefore, the general properties of the matrices P and Q were investigated and equivalent expressions found that contain only known quantities. The statistics are given for s hypothetical functions. The details of derivation are given in the papers

(a) "Direction and Dimensionality tests based on Hotelling's generalized T_o^2" by A. M. Kshirsagar [accepted for publication in "Perspectives in Probability and Statistics" edited by J. Gani (papers in honour of M. S. Bartlett), published by Applied Probability Trust, Sheffield, England].

(b) "Testing the adequacy of s hypothetical discriminant functions using Pillai's V and Roy's largest root criterion" by McHenry (to appear in the Journal of the Indian Statistical Association, Poona, India).

Copies of these papers are appended.

The case when the hypothetical function or functions are given from the space of dummy variables, is also considered. Kshirsagar [1971] has succeeded in expressing the factors of Wilks' Λ as the product of independent Beta variables by making a series of matrix transformations. He constructs a matrix M which has a central matrix variate Beta distribution under the null hypothesis and shows that the direction and collinearity statistics are related to submatrices of this M. Mitra [1970] has studied the matrix variate beta distribution thoroughly. By using some of his results in conjunction with Kshirsagar's results it is shown that Hotelling's generalized T_o^2 and Pillai's V criteria can be constructed for testing the adequacy of a hypothetical discriminant function from the $\underline{y}$-space also. Again, the test statistics are given for s assigned functions. Roy's maximum root criterion is also considered for this purpose.

The details are given in the following two papers:

(a) "Use of Hotelling's generalized T_o^2 in multivariate tests" by C. E. McHenry and A. M. Kshirsagar. (To appear in Multivariate Analysis-IV (Symposium on Multivariate Analysis at Dayton, Ohio, 1975), published by North-Holland Publishing Co.)

(b) "Tests of Hypothetical contrasts in Manova using two different multivariate criteria", by A. M. Kshirsagar and C. E. McHenry. (To appear in "Report of Statistical Applications Research", Japan.)

Copies of these papers are appended.

2. APPLICATIONS

$\underline{x}$-Space

The exact percentage points of Wilks' Λ, Hotelling's T_o^2 and Pillai's V have been tabulated but in practice they are rarely needed as the χ^2 or F approximations to their distributions are very satisfactory. In this report the χ^2 approximation suggested by Bartlett [1938] will be used for Wilks' Λ and the F approximations suggested by Pillai and Mijares [1959] and Pillai and Samson [1959] will be used for Pillai's V and Hotelling's T_o^2.

Consider the example of G. I. Taylor's blood-group seriological data analyzed by Bartlett [1951]. He used only Wilks' criterion for the analysis. He also remarks that these non-numerical data would hardly justify an elaborate analysis but they constitute a very convenient and interesting example for purposes of illustrating possible tests. This is true because precise tests of significance are based on the assumption of normality for at least one set of variables, an assumption which cannot strictly be true in this example. If the tests are correctly formulated, they will be asymptotically correct, and therefore, still informative.

The data consist of 144 reactions obtained by testing twelve samples of human blood with twelve different sera, these reactions being represented by the symbols –, ?, w, (+) and +. From the point of view of multivariate analysis, the coefficients of the discriminant functions adequate for discriminating among these reactions are the coefficients of the four dummy variables, x_1, x_2, x_3 and x_4, which are given the following values:

Table 1

	-	?	w	(+)	+
x_1	0	1	0	0	0
x_2	0	0	1	0	0
x_3	0	0	0	1	0
x_4	0	0	0	0	1

The matrices of the sum of squares and sum of products for these four variables are quoted from Bartlett and given in Tables 2 and 3.

Table 2

Total s.s. and s.p., with n = 143 d.f.

	x_1	x_2	x_3	x_4
x_1	1703	-1157	-468	-65
x_2	-1157	4895	-3208	-445
x_3	-468	-3205	3888	-180
x_4	-65	-445	-190	695

Table 3

Between s.s. and s.p., with q = 22 d.f.

	x_1	x_2	x_3	x_4
x_1	718	2	-672	-106
x_2	2	1630	-1416	-218
x_3	-672	-1416	1944	216
x_4	-106	-218	216	118

We now come to the question of testing the adequacy of a hypothetical discriminant function, viz.

$$\underline{\ell}'\underline{x} = .25x_1 + .5x_2 + .75x_3 + x_4$$

for discriminating among reactions. The numerical values of the tests of direction and 'partial' collinearity are given in Table 4 and the alternative factorization is given in Table 5. The letter F is used to denote the approximate F distributional value and the corresponding probability is that given by the approximation of getting a larger value.

The numerical results for all the three criteria imply that the number of specified functions is correct but that the assigned discriminant function is inadequate for discrimination. This is true because of the insignificance of the collinearity factor and

TABLE 4

Taylor's blood-group data

Test	Approximation	Probability
$\Lambda_D = .916$	$\chi^2_3 = 12.226$	.0071
$T^2_D = .0916$	$F_{3,139} = 4.244$	.0069
$V_D = .0839$	$F_{3,139} = 4.244$	.0069
$\Lambda_{C\|D} = .565$	$\chi^2_{63} = 73.439$	.182
$T^2_{C\|D} = .644$	$F_{63,350} = 1.193$	.165
$V_{C\|D} = .511$	$F_{63,360} = 1.172$	.189

TABLE 5

Alternative factorization for Taylor's blood-group data

Test	Approximation	Probability
$\Lambda_C = .572$	$\chi^2_{63} = 72.344$	.206
$T^2_C = .630$	$F_{63,353} = 1.176$	.185
$V_C = .499$	$F_{63,363} = 1.151$	.216
$\Lambda_{D\|C} = .904$	$\chi^2_3 = 11.911$	.0082
$T^2_{D\|C} = .106$	$F_{3,118} = 4.159$	.0079
$V_{D\|C} = .0956$	$F_{3,118} = 4,159$	.0079

the significance of the direction factor.

To illustrate the tests further, consider the two hypothetical functions

$$\underline{\ell}_1'\underline{x} = .25x_1 + .5x_2 + .75x_3 + x_4$$

and

$$\underline{\ell}_2'\underline{x} = x_1 + 3x_2 + 2x_3 + 2x_4$$

and test the hypothesis of their adequacy for discrimination. The resulting numerical values are given in Tables 6 and 7.

TABLE 6

Taylor's blood-group data (two functions)

Test	Approximations	Probability
$\Lambda_D = .917$	$\chi^2_4 = 12.045$	.0173
$T^2_D = .0909$	$F_{4,274} = 3.112$	.0157
$V_D = .0833$	$F_{4,278} = 3.020$	.0182
$\Lambda_{C\|D} = .775$	$\chi^2_{40} = 32.573$	.789
$T^2_{C\|D} = .273$	$F_{40,234} = .798$	.802
$V_{C\|D} = .240$	$F_{40,238} = .810$	.785

TABLE 7

Alternative factorization for Taylor's blood-group data

(two functions)

Test	Approximation	Probability
$\Lambda_C = .784$	$\chi^2_{40} = 31.486$	.826
$T^2_C = .259$	$F_{40,238} = .769$	.839
$V_C = .229$	$F_{40,242} = .782$	.824
$\Lambda_{D\mid C} = .905$	$\chi^2_4 = 11.767$	.019
$T^2_{D\mid C} = .104$	$F_{4,234} = 3.054$	.018
$V_{D\mid C} = .0945$	$F_{4,238} = 2.952$	.021

Again all the test criteria agree that the number of functions specified is adequate but the hypothetical functions given are not. But we have seen earlier that only one function was needed so two will certainly be adequate. Therefore if the direction factor had not been significant, we would be faced with the problem of choosing the optimum subset of hypothetical functions that discriminate adequately. This is just the problem of variable selection and will not be discussed here.

This example does not imply that the three criteria give contradictory results; therefore, until further investigations are made about their power, the choice is merely a matter of personal taste and prejudices.

$\underline{y}$-Space

It was shown earlier that the hypothetical functions given in the dummy variables' space may be either in the form of contrasts among the k population means $\underline{\mu}_i$ (i = 1, ..., k) or as non-stochastic and exogeneous variables such as time. Examples will be given illustrating both.

Consider the anthropometric data consisting of observations on four series of Egyptian skulls, first investigated by Barnard [1935], and later analyzed further by Bartlett [1947], Rao [1952], Williams [1959], and Kshirsagar [1962b] in which time, t, was to be judged for its adequacy as a discriminator.

The four measurements, basialveolar length x_1, nasal height x_2, maximum breadth x_3, and basibregmatic height x_4 were taken and the relevant sum of squares and sum of products are given in Tables 8 and 9.

The corrected sum of products of $\underline{x}$ with t is

$$C_{xt} = \begin{vmatrix} 718.7628 \\ -1407.2608 \\ 410.1019 \\ -733.4276 \end{vmatrix}$$

and the total sum of squares of t is $C_{tt} = 4307.6683$.

TABLE 8

Between s.s. and s.p., with q = 3 d.f.

	x_1	x_2	x_3	x_4
x_1	123.1806	-231.3756	87.3053	-128.7640
x_2	-231.3756	486.3459	-107.5056	125.3133
x_3	87.3053	-107.5056	150.4115	-137.5808
x_4	-128.7640	125.3133	-137.5808	640.7339

TABLE 9

Within s.s. and s.p., with n - q = 394 d.f.

	x_1	x_2	x_3	x_4
x_1	9661.9975	445.5733	1130.6239	2148.5842
x_2	445.5733	9073.1150	1239.2220	2255.8127
x_3	1130.6239	1239.2220	3928.3203	1271.0547
x_4	2148.5842	2255.8127	1271.9547	8741.5088

The results of Wilks' Λ, Hotelling's T_o^2 and Pillai's V for testing the adequacy of t as a discriminator are given in Tables 10 and 11. All the three criteria show that t can be used as a discriminator but it is not sufficient by itself because of the significance of

TABLE 10

Egyptian skull data

Tests	Approximation	Probability
$*\Lambda_D = .989$	$\chi^2_2 = 4.403$	.109
$*T^2_D = .0112$	$F_{2,394} = 2.214$	.108
$*V^2_D = .0111$	$F_{2,394} = 2.214$	.108
$*\Lambda_{C\|D} = .909$	$\chi^2_6 = 37.360$	.000
$*T^2_{C\|D} = .0984$	$F_{6,780} = 6.397$	.000
$*V_{C\|D} = .0923$	$F_{6,784} = 6.325$	.000

TABLE 11

Alternative factorization for Egyptian skull data

Tests	Approximation	Probability
$*\Lambda_C = .9097$	$\chi^2_6 = 37.201$	.000
$*T^2_C = .0977$	$F_{6,782} = 6.368$	.000
$*V_C = .0917$	$F_{6,786} = 6.298$	.000
$*\Lambda_{D\|C} = .988$	$\chi^2_2 = 4.623$	.0975
$*T^2_{D\|C} = .0119$	$F_{2,391} = 2.325$	.0970
$*V_{D\|C} = .0118$	$F_{2,391} = 2.325$	.0970

the dimensionality factor.

Williams [1967] analyzed some numerical data extracted from the results of a study of the relation of lamb carcasses (Robinson et al. [1956]) using Wilks' criterion. The analysis will now be given using Hotelling's T_o^2 and Pillai's V.

Since grade was subjectively determined it was believed that a more consistent assessment could be made if grade was found to be correlated with the measurements of the carcass. On the other hand, if it was not found to be highly correlated with such measurements, one might question the methods of grading.

A selection of the original data, comprising results for 20 carcasses in each of three grades and two weight-classes, is presented in Table 12 and quoted from Williams. The four measurements, $\underline{x}$, whose association with grade is sought are width of shoulder x_1, thickness of flank x_2, width of flank x_3 and length of leg x_4. The three grades are Down Royal (or briefly, Down), Royal and Tallarook. Between the combinations of these grades and the two weight-classes are five comparisons, described by y_1, y_2, y_3, y_4 and y_5; for convenience these are defined to represent weight differences, equal spacing of grades, departure from equal spacing and the interactions of these two comparisons with the weight-classes. These are set out in the final columns of Table 12. The original data are not presented, but are summarized in the product matrices shown in Tables 13, 14, 15 and 16, and quoted from Williams.

TABLE 12

Average values of characteristics of lamb carcasses

Weight-Class	Grade	Measurements x_1	x_2	x_3	x_4	Categories y_1	y_2	y_3	y_4	y_5
Medium	Down	189	16.0	222	223	-1	-1	+1	+1	-1
	Royal	181	14.8	218	257	-1	0	-2	0	+2
	Tallarook	178	12.2	219	263	-1	+1	+1	-1	-1
Heavy	Down	208	19.6	236	232	+1	-1	+1	-1	+1
	Royal	196	16.4	236	262	+1	0	-2	0	-2
	Tallarook	196	14.6	232	268	+1	+1	+1	+1	+1

TABLE 13

Within s.s. and s.p., with n - q = 114 d.f.

	x_1	x_2	x_3	x_4
x_1	6939	286	2330	-2680
x_2	286	558	189	-199
x_3	2330	189	12194	-3280
x_4	-2680	-199	-3280	12260

TABLE 14

Between s.s. and s.p. with q = 5 d.f.

	x_1	x_2	x_3	x_4
x_1	11836	2351	8376	-6759
x_2	2351	611	1437	-2793
x_3	8376	1437	7335	381
x_4	-6759	-2793	381	34254

Table 15

C_{xy} s.p. of $\underline{x}$ and $\underline{y}$

	y_1	y_2	y_3	y_4	y_5
x_1	1020	-453	339	7	171
x_2	155	-177	-3	-25	53
x_3	914	-132	-2	-20	-190
x_4	379	1505	-1059	-67	61

TABLE 16

C_{yy} s.s. and s.p. of $\underline{y}$

	y_1	y_2	y_3	y_4	y_5
y_1	120	0	0	0	0
y_2	0	80	0	0	0
y_3	0	0	240	0	0
y_4	0	0	0	80	0
y_5	0	0	0	0	240

The results of each test criterion for testing the adequacy of y_1, y_2, and y_3 as discriminators are given in Tables 17 and 18. Again, all three criteria agree that y_1, y_2, and y_3 are adequate as discriminators.

From the examples considered in sections 4.1 and 4.2, Wilks' Λ, Hotelling's T_o^2 and Pillai's V seem to perform equally well but obviously nothing can be said in general until in depth study of their power is made.

TABLE 17

Lamb carcass data

Tests	Approximation	Probability
$*\Lambda_D = .963$	$\chi^2_6 = 4.286$	.640
$*T^2_D = .0384$	$F_{6,222} = .711$	.643
$*V_D = .0374$	$F_{6,226} = .718$	.637
$*\Lambda_{C\|D} = .960$	$\chi^2_2 = 4.485$	.105
$*T^2_{C\|D} = .0412$	$F_{2,111} = 2.288$	.104
$*V_{C\|D} = .0396$	$F_{2,111} = 2.288$	.104

TABLE 18

Alternative factorization for lamb carcass data

Tests	Approximation	Probability
$*\Lambda_C = .961$	$\chi^2_2 = 4.517$	.103
$*T^2_C = .0404$	$F_{2,114} = 2.304$	.102
$*V_C = .0389$	$F_{2,114} = 2.304$	.102
$*\Lambda_{D\|C} = .962$	$\chi^2_6 = 4.335$	.634
$*T^2_{D\|C} = .0392$	$F_{6,220} = .719$	.636
$*V_{D\|C} = .0382$	$F_{6,224} = .727$	.631

3. ANALYSIS OF QUALITATIVE DATA

Let A and B be two attributes with p + 1 categories $A_1, \ldots, A_{p+1}$ for A and q+1 categories $B_1, B_2, \ldots, B_{q+1}$ for B. Out of a sample of n items, let us assume that n_{ij} items ($i = 1, \ldots, p+1$; $j = 1, \ldots, q+1$) belong to the cell (A_i, B_j) i.e. they are in the category A_i and also in the category B_j. This will give the following contingency table, with marginal totals $n_{i.}$ and $n_{.j}$

Contingency Table

A \ B	B_1	...	B_j	...	B_{q+1}	Totals
A_1	n_{11}	...	n_{1j}	...	n_{1q+1}	$n_{1.}$
⋮						
A_i	n_{i1}	...	n_{ij}	...	n_{iq+1}	$n_{i.}$
⋮						
A_{p+1}	$n_{p+1,1}$	...	$n_{p+1,j}$	...	$n_{p+1,q+1}$	n_{p+1}
Totals	$n_{.1}$	...	$n_{.j}$	...	$n_{.q+1}$	n

Let π_{ij} be the probability that an item belongs to the (A_i, B_j) cell ($i = 1, \ldots, p+1$; $j = 1, \ldots, q+1$). $\sum_i \sum_j \pi_{ij} = 1$. If the two attributes are independent

$$\pi_{ij} = \text{Prob(that the item belongs to } A_i) \cdot \text{Prob(it belongs to } B_j)$$

$$= \pi_{i.} \, \pi_{.j}$$

where

$$\sum_i \pi_{ij} = \pi_{ij} \; , \quad \sum_j \pi_{ij} = \pi_{i.}$$

The matrix

$$\pi = (\pi_{ij} - \pi_{i.} \, \pi_{.j})$$

is of rank 0. Usually no further analysis is done beyond testing whether A and B are independent or not. If they are not independent, it is instructive to investigate the nature of the dependence. This problem can be formally brought under the theory of canonical analysis by defining

$$x_i = \begin{cases} 1 & \text{if an item belongs to } A_i \\ 0 & \text{otherwise} \end{cases} \qquad i = 1, \ldots, p+1$$

and

$$y_j = \begin{cases} 1 & \text{if an item belongs to } B_j \\ 0 & \text{otherwise} \end{cases} \qquad j = 1, \ldots, q+1.$$

$\underline{x}' = [x_1, \ldots, x_{p+1}]$, $\underline{y}' = [y_1, \ldots, y_{q+1}]$. The covariance matrix of $\underline{x}$ and $\underline{y}$ can be easily seen to be π defined earlier. If it is not null, the next simplest structure is

$$\pi = \lambda \, \underline{\alpha} \, \underline{\beta}'$$

i.e. it is of rank 1 and $\underline{\alpha}$, $\underline{\beta}$ are its left and right eigenvectors. If so, it can be readily verified that $\underline{\alpha}'\underline{x}$ and $\underline{\beta}'\underline{y}$ are the only linear functions of $\underline{x}$ and $\underline{y}$ that are correlated and that any other linear functions are uncorrelated. Then it is obvious that elements of $\underline{\alpha}$ are the optimum scores for $A_1, \ldots, A_{p+1}$ and elements of $\underline{\beta}$ are the optimum scores for $B_1, \ldots, B_{q+1}$. In the terminology of canonical analysis, $\underline{\alpha}'\underline{x}$, $\underline{\beta}'\underline{y}$ are canonical variables and these are the only ones having a non-null canonical correlation. If therefore, a set of scores h_1, h_2, ..., h_{p+1} are proposed for the categories of A, one can test the goodness of fit of these scores by exactly the same procedures as in the test of goodness of fit of a hypothetical discriminant function.

All the test statistics Λ, T_o^2, V, r_1^2 can be employed, as discussed earlier. They can be further factorized also to investigate the reason for departure from goodness of fit whether the proposed scores are not the right ones or whether only one set of scores is not adequate (this will be so if the matrix π is of rank 2 or more) or both.

It is however desirable that these statistics now be expressed in terms of the cell frequencies n_{ij}. This is accomplished in this project and the details of these results will be separately published in a research paper which is now under preparation.

4. POSSIBLE APPLICATIONS OF THESE RESEARCH RESULTS TO USAF PROBLEM AREAS

This research work is likely to be of use in constructing suitable "indices", to describe operational performance of pilots, navigators and other personnel from the qualitative data collected for this purpose. These indices will be "composite" functions of the optimum scores attached to different categories of their performance and behavior and can be constructed from the canonical variables. This will help in comparing the operational efficiency of one group with another, or of the same group over different periods of time or different operational conditions. The effectiveness of a training program to improve the efficiency can also be assessed from such scores and indices. Comparison of alternative programs can also be done by using these indices. This will help, therefore, in optimal decisions about choice of a suitable group or of a suitable program, for any activity.

5. MULTIVARIATE RELIABILITY

In this section one of the most basic stochastic models in reliability theory, namely, the on-off process generated by failures and repairs of components in a series will be considered. Consider a series system of p components that operates if and only if each of the p components operates. None of the components operate while the system is down and only failed components are repaired or replaced. Furthermore, replacement is assumed to take a random time and repaired components are assumed to function like new ones.

The case of availability, the probability that the system is functioning, is treated extensively in the literature. But most papers assume special repair or failure distributions (or both). A more general situation will now be discussed.

Let X_{ir} be the length of the r^{th} functioning period of the i^{th} component with distribution F_i, assumed continuous, and mean μ_i, $i = 1, 2, \ldots, p$ (i.e., time to failure of the r^{th} replacement for component i excluding down times for the system). Further, let D_{ir} be the length of r^{th} repair time for the i^{th} component and has the distribution G_i with mean v_i, $i = 1, 2, \ldots, p$. It is also assumed that $\{X_{ir}\}_{r=1}^{\infty}$ and $\{D_{ir}\}_{r=1}^{\infty}$ are mutually independent renewal processes. Barlow and Proschan [1973] have given a more in depth discussion of this process.

If we define $\delta(t) = i$ if the system is down at time t due to the failure of the i^{th} component and $\delta(t) = 0$ if the i^{th} component is functioning at time t, then the limiting average availability is

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t p[\delta(u) = 0]du = \left[1 + \sum_{j=1}^{p} \frac{v_j}{\mu_j}\right]^{-1} = \pi_0 ,$$

while

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t p[\delta(u) = i]du = \frac{v_i}{\mu_i} \pi_0 = \pi_i ,$$

where $i = 1, 2, \ldots, p$. These results are true for arbitrary failure and repair distributions.

Now, define $N_i(t)$ to be the number of replacements of component i in time t and it follows that the average number of replacements for component i is

$$\lim_{t\to\infty} \frac{E[N_i(t)]}{t} = \frac{\pi_0}{\mu_i} = m_i^{-1} \qquad i = 1, \ldots, p .$$

Thus, approximately tm_i^{-1} will be the number of spare parts required for the i^{th} component in $[0, t]$. m_i is the mean time between failures of the i^{th} component.

It has been shown (see Barlow and Proschan [1973]) that if we define $\sigma_i^2 = Var(X_i)$, $\tau_i^2 = Var(D_i)$ and

$$N_i^*(t) = t^{-1}[N_i(t) - tm_i^{-1}]$$

that under rather general conditions

$$(N_1^*(t), N_2^*(t), \ldots, N_p^*(t))$$

has an asymptotically $(t \to \infty)$ multivariate normal distribution with mean vector $\underline{0}$ and variance-covariance matrix

$$\Sigma = (\sigma_{ij})$$

where

$$\sigma_{ij} = m_i^{-1} m_j^{-1} \{ \pi_0 \sum_{s=1}^{p} [v_s^2 \sigma_s^2 \mu_s^{-3} + \tau_s^2 \mu_s^{-1}]$$

$$- v_i \sigma_i^2 \mu_i^{-2} - v_j \sigma_j^2 \mu_j^{-2} \} \qquad (i \neq j) ,$$

$$\sigma_{ii} = v_i^2 = m_i^{-3} w_i^2 ,$$

$$w_i^2 = \sigma_i^2 c_i^2 + \mu_i \sum_{j=1}^{p} \tau_j^2 \mu_j^{-1} + \mu_i \sum_{j \neq i} \sigma_j^2 v_j \mu_j^{-3} ,$$

and

$$c_i = 1 + \sum_{j \neq i} v_j \mu_j^{-1} , \quad \pi_0 = \left[1 + \sum_{j=1}^{p} v_j \mu_j^{-1} \right]^{-1} .$$

Therefore, if k systems each consisting of p components are to be investigated with respect to $N_i(t)$ and t is large then the methods and techniques derived in this research may be used. For example, to investigate the mean time between failures for the k systems, in order to determine the optimum, contrasts in the $\underline{y}$-space (i.e., the space generated by the different means m_i) can be specified and tested by the criteria derived in this report. If the relationship among the $N_i(t)$ is to be investigated then hypothetical functions from the $\underline{x}$-space could be used and the results in the appended papers applied.

This system of p components can be generalized to a Markov renewal process. That is, assume that when the $(i - 1)^{th}$ component fails it can be replaced by any of k different components which function in the same capacity as the one that failed. The distribution of the replacement which is to function during the i^{th} period is F_{ih} $(h = 1, 2, \ldots, z)$, F_{ih} is not necessarily the same distribution as $F_{ih'}$ $(h \neq h')$. Furthermore, the choice of F_{ih} satisfies a Markov chain. In other words, if the $(i - 1)^{th}$ component, with distribution $F_{ih'}$, fails then the probability that the replacement has the distribution F_{ih} is $p(h', h)$ (i.e., a transition probability in a Markov chain). Again, it is assumed that X_{ir} and D_{ir} $(i = 1, 2, \ldots, p;\ r = 1, 2, \ldots)$ are mutually independent processes where D_{ir} is the duration of the repair time for the i^{th} component and has the distribution G_i. Further research is needed in this area.

6. CONCLUSIONS AND FUTURE RESEARCH

Test statistics of Hotelling's T_o^2, Pillai's V and Roy's r_1^2 type corresponding to the factorizations of Wilks' Λ have been derived for testing the adequacy of s hypothetical discriminant functions. The proposed functions may either be given in the $\underline{x}$-space or the dummy variables' space. If the assigned functions are given in the $\underline{x}$-space and $q < p$, the distributions of each of the factorizations for Wilks' Λ, Pillai's V, Hotelling's T_o^2 and Roy's maximum root criterion are derived. The case of $q \geq p$ is covered in the literature (see Kshirsagar [1970]). The distributions of each of the criteria are also derived when the s proposed functions are given in the $\underline{y}$-space and $q - s < p$ by using the properties of the matrix variate beta distribution. Again, the distributions for $q - s \geq p$ can be found in the literature (see Kshirsagar [1971]).

For the illustrative examples presented here, V, T_o^2 and Λ performed equally well as test criteria. Roy's maximum root criterion was not used in any of the examples. These results were not surprising since investigations comparing the powers of Λ, T_o^2, V and r_1^2 using Monte Carlo techniques have not resulted in any general significant differences. Now that these test statistics have been spelled out for all four criteria maybe the investigations into their power when testing the adequacy of s hypothetical discriminant functions will yield some fruitful results.

It is the belief of the author that Wilks' Λ is superior to the three other criteria because the original factorizations were given for Λ and no such partitioning seems to originate in T_o^2, V or r_1^2. Further, the expressions for the factorizations of Λ are much simpler than those of the other criteria.

In multivariate analysis of variance, one still finds statisticians contenting themselves with only overall tests of significance instead of exploring the structure of the relations in greater detail. For this reason, computer programs implementing the tests given in this report have been written. The results in chapter four were obtained by utilizing these programs.

Future Research

One of the main problems in multivariate analysis is the selection of variables. Methods have been suggested for selecting variables but most are not feasible and do not use all the information available in the data. The use of direction and collinearity factors as a selection procedure should be investigated. The collinearity factor could be used to indicate a sufficient number of variables while the direction factor would indicate the optimum variables to be used for that subset size.

Another area open to research is in the case of singular data. That is, if multicollinearity exists among the observation variables, what effect does it have on each of the multivariate test criteria? Is it possible to substitute the generalized inverses where the matrix inverses are normally used? And are there any restrictions on the generalized inverse to be used?

Research in the area of missing data is also badly needed. Oftentimes the multivariate data is incomplete due to reasons beyond the control of the experimenter. Rao [1965] has suggested a procedure based on Wilks' Λ for such situations. Factorize Λ as

$$\Lambda_1 \Lambda_2 \cdots \Lambda_p$$

where Λ_1 is based on x_1 alone, Λ_2 on x_2 eliminating x_1 and so on. If x_p is the variable, on which r of the observations are not recorded, Rao suggests using

$$\Lambda_1 \Lambda_2 \cdots \Lambda_p^{1/r}$$

as a test criterion, by discounting Λ_p proportionately. Not much has been done in this area.

Another field open for further study is testing the homogeneity of discriminant functions. Suppose it is desired to use the same discriminant functions for evaluating the social structures of two different countries, can one use the same set of discriminant functions for both countries" How does one construct a test in this situation?

Lastly, if the number of proposed functions is over specified, how can it be detected and dealt with?

REFERENCES

Barlow, R. E. and Prochan, F. [1973]. Availability theory for multicomponent systems. Multivariate Analysis - III, edited by P. R. Krishnaiah, Academic Press, New York, 319-35.

Barnard, M. M. [1935]. The secular variation of skull characters in four series of Egyptian skulls. Ann. Eugen. 7, 189-200.

Bartlett, M. S. [1938]. Further aspects of the theory of multiple regression. Proc. Camb. Phil. Soc. 34, 33-45.

Bartlett, M. S. [1939]. The standard errors of discriminant function coefficients. J. R. Statist. Soc. B 6, 169-73.

Bartlett, M. S. [1947]. Multivariate analysis. J. R. Statist. Soc. B 9, 176-90.

Bartlett, M. S. [1948]. A note on the statistical estimation of demand and supply relations from time series. Econometrica 16, 323-30.

Bartlett, M. S. [1951]. The goodness of fit of a single hypothetical discriminant function in the case of several groups. Ann. Eugen. 16, 199-214.

Deemer, W. L. and Olkin, I. [1951]. The Jocabians of certain matrix transformations useful in multivariate analysis. Biometrika 38, 345-67.

Fisher, R. A. [1938]. The statistical utilization of multiple measurements. Ann. Eugen. 8, 376-86.

Fisher, R. A. [1940]. The precision of discriminant functions. Ann. Eugen. 10, 422-29.

Good, I. J. [1969]. Some applications of the singular decomposition of a matrix. Technometrics 11, 823-31.

Hotelling, H. [1951]. A generalized T test and measure of multivariate dispersion. Proc. 2nd Berkeley Symp., 23-41.

Ito, K. [1960]. On multivariate analysis of variance tests. Bull. Int. Inst. Statist. 38, 88-100.

Kshirsagar, A. M. [1959]. Bartlett decomposition and Wishart distribution. Ann. Math. Statist. 30, 239-41.

Kshirsagar, A. M. [1960]. A note on the derivation of some exact multivariate tests. Biometrika 47, 480-2.

Kshirsagar, A. M. [1961]. The non-central multivariate beta distribution. Ann. Math. Statist. 32, 104-11.

Kshirsagar, A. M. [1962a]. Prediction from the simultaneous equations system and Wold's implicit causual chain model. Econometrica 30, 801-11.

Kshirsagar, A. M. [1962b]. A note on direction and collinearity factors in canonical analysis. Biometrika 49, 255-59.

Kshirsagar, A. M. [1963a]. Confidence intervals for discriminant function coefficients. J. Ind. Statist. Assoc. 1, 1-8.

Kshirsagar, A. M. [1963b]. Effect of non-centrality on the Bartlett decomposition of a Wishart matrix. Ann. Inst. Statist. Math. 14, 217-28.

Kshirsagar, A. M. [1964a]. Distributions of the direction and collinearity factors in discriminant analysis. Proc. Camb. Phil. Soc. 60, 217-25.

Kshirsagar, A. M. [1964b]. Wilks' Λ. J. Ind. Statist. Assoc. 2, 1-20.

Kshirsagar, A. M. [1965]. A note on the use of canonical analysis in factorial experiments. J. Ind. Statist. Assoc. 3, 165-69.

Kshirsagar, A. M. [1969a]. Distributions associated with the factors of Wilks' in discriminant analysis. J. Aust. Math. Soc. 10, 269-77.

Kshirsagar, A. M. [1969b]. Correlation between two vector variables. J. R. Statist. Soc. B 31, 477-85.

Kshirsagar, A. M. [1970]. An alternative derivation of the distribution of the direction and collinearity statistics in discriminant analysis. Cal. Statist. Assoc. Bull. 19, 123-34.

Kshirsagar, A. M. [1971]. Goodness of fit of a discriminant function from the vector space of dummy variables. J. R. Statist. Soc. B 33, 111-16.

Kshirsagar, A. M. [1972]. Multivariate Analysis. Marcel Dekker, Inc., New York.

Mikhail, M. N. [1965]. A comparison of tests of the Wilks-Lawley hypothesis in multivariate analysis. Biometrika 52, 149-56.

Mitra, S. K. [1970]. A density-free approach to the matrix variate beta distribution. Sankhya A 32, 81-88.

Pillai, K. C. S. [1955]. Some new test criteria in multivariate analysis. Ann. Math. Statist. 26, 117-21.

Pillai, K. C. S. and Mijares, T. A. [1959]. On the moments of the trace of a matrix and approximations to its distribution. Ann. Math. Statist. 30, 1135-40.

Pillai, K. C. S. and Samson, Jr., P. [1959]. On Hotelling's generalization of T_0^2 Biometrika 46, 160-5.

Pillai, K. C. S. and Dotson, C. O. [1969]. Power comparisons of tests of two multivariate hypothesis based on individual characteristic roots. Ann. Inst. Statist. Math. 21, 49-60.

Pillai, K. C. S. and Jayachandran, K. [1967]. Power comparisons of tests of two multivariate hypothesis based on four criteria. Biometrika 54, 195-210.

Pillai, K. C. S. and Jayachandran, K. [1968]. Power comparisons of tests of equality of two covariance matrices based on four criteria. Biometrika 55, 335-42.

Rao, C. R. [1952]. Advanced Statistical Methods in Biometric Research. Wiley, New York.

Robinson, T. J., Binet, F. E. and Doig, Alison G. [1956]. An assessment of the relative value of various external measurements for differentiating between various grades of export lamb carcasses. Aust. J. Agric. Res. 7, 345-65.

Roy, S. N. [1957]. Some Aspects of Multivariate Analysis. Wiley, New York.

Wilks, S. S. [1932]. Certain generalizations of the analysis of variance. Biometrika 24, 471-81.

Williams, E. J. [1952a]. Some exact tests in multivariate analysis. Biometrika 39, 17-31.

Williams, E. J. [1952b]. Use of scores for the analysis of association in contingency tables. Biometrika 39, 274-89.

Williams, E. J. [1952c]. The interpretation of interactions in factorial experiments. Biometrika 39, 65-81.

Williams, E. J. [1955]. Significance tests for discriminant functions and linear functional relationships. Biometrika 42, 360-81.

Williams, E. J. [1958]. Simultaneous regression equations in experimentation. Biometrika 45, 96-110.

Williams, E. J. [1959]. Regression Analysis. Wiley, New York.

Williams, E. J. [1961]. Tests for discriminant functions. J. Aust. Math. Soc. 2, 243-52.

Williams, E. J. [1967]. The analysis of association among many variates. J. R. Statist. Soc. B 29, 199-242.

APPENDIX 1

(x-space)

Direction and dimensionality factors and their distributions

Factor	Distribution
$\Lambda_D = \lvert L_1B_x(W_x + B_x)^{-1}W_xL_1'\rvert / \lvert L_1B_xL_1'\rvert \Lambda_H$	$\Lambda(n - s, p - s, s)$
$\Lambda_{C\lvert D} = \Lambda_R/\Lambda_D$	$\Lambda(n - 2s, p - s, q - s)$
$T_D^2 = \mathrm{tr}\ L_1B_xL_1'(L_1B_xL_1' + L_1W_xL_1')^{-1}L_1W_xL_1'[L_1W_x(W_x + B_x)^{-1}B_xL_1']^{-1} - s$	$T_o^2(n - s, p - s, s)$
$T_{C\lvert D}^2 = \mathrm{tr}\ W_x^{-1}B_x - \mathrm{tr}\ L_1B_xW_x^{-1}B_xL_1'(L_1B_xL_1')^{-1}$	$T_o^2(n - 2s, p - s, q - s)$
$V_D = s - \mathrm{tr}[(L_1B_xL_1')^{-1} + (L_1W_xL_1')^{-1}]L_1W_x(W_x + B_x)^{-1}B_xL_1'$	$V(n - s, p - s, s)$
$V_{C\lvert D} = \mathrm{tr}(W_x + B_x)^{-1}B_x - \mathrm{tr}\ L_1B_x(W_x + B_x)^{-1}W_x(W_x + B_x)^{-1}B_xL_1'[L_1W_x$	
$\cdot\ (W_x + B_x)^{-1}B_xL_1']^{-1}$	$V(n - 2s, p - s, q - s)$
$r_D^2 = \max_{r^2}\{r^2 : \lvert -r^2I + (W_x + B_x)^{-1}F_xF_x'\rvert = 0\}$	$r_1^2(n - s, p - s, s)$
$r_{C\lvert D}^2 = \max_{r^2}\{r^2 : \lvert -r^2I + [W_x + B_x - B_xL_1'(L_1B_xL_1')^{-1}L_1B_x]^{-1}[B_x$	
$- B_xL_1'(L_1B_xL_1')^{-1}L_1B_x]\rvert = 0\}$	$r_1^2(n - 2s, p - s, q - s)$

APPENDIX 2

($\underline{x}$-space)

Alternative factorization of Λ, T_o^2, V and r_1^2

Factor	Distribution
$\Lambda_C = \Lambda \lvert L_1B_xL_1' + L_1B_xW_x^{-1}B_xL_1' \rvert / \lvert L_1B_xL_1' \rvert$	$\Lambda(n - s, q - s, p - s)$
$\Lambda_{D\mid C} = \Lambda_R/\Lambda_C$	$\Lambda(n - q, s, p - s)$
$T_C^2 = \mathrm{tr}\, W_x^{-1}B_x - \mathrm{tr}\, L_1B_xW_x^{-1}(W_x + B_x)W_x^{-1}B_xL_1'(L_1B_xL_1' + L_1B_xW_x^{-1}B_xL_1')^{-1}$	$T_o^2(n - s, q - s, p - s)$
$T_{D\mid C}^2 = \mathrm{tr}(L_1B_xL_1')^{-1}L_1B_xW_x^{-1}B_xL_1'(L_1B_xL_1' + L_1W_xL_1')^{-1}L_1W_xL_1' - \mathrm{tr}\, L_1B_xL_1' \cdot (L_1B_xL_1' + L_1W_xL_1')^{-1}$	$T_o^2(n - q, s, p - s)$
$V_C = \mathrm{tr}(W_x + B_x)^{-1}B_x + \mathrm{tr}\, L_1B_x(W_x + B_x)^{-1}W_xL_1'(L_1B_xL_1')^{-1} - s$	$V(n - s, q - s, p - s)$
$V_{D\mid C} = \mathrm{tr}[L_1B_xW_x^{-1}B_xL_1' - L_1B_xL_1'(L_1W_xL_1')^{-1}L_1B_xL_1'](L_1B_xL_1' + L_1B_xW_x^{-1}B_xL_1')^{-1}$	$V(n - q, s, p - s)$
$r_C^2 = \max\limits_{r^2}\{r^2 : \lvert -r^2I + (W_x + B_x)^{-1}[B_x - B_xL_1'(L_1B_xL_1')^{-1}L_1B_x] \rvert = 0\}$	$r_1^2(n - s, q - s, p - s)$
$r_{D\mid C}^2 = \max\limits_{r^2}\{r^2 : \lvert -r^2I + [W_x + B_xL_1'(L_1B_xL_1')^{-1}L_1B_x]^{-1}F_xF_x' \rvert = 0\}$	$r_1^2(n - q, s, p - s)$

APPENDIX 3

($\underline{y}$-space)

Direction and dimensionality factors and their null distributions

Factor	Distribution
$*\Lambda_D = \lvert C_{tx}(W_x + L_x)^{-1}W_x(W_x + L_x)^{-1}C_{xt}\rvert / \lvert C_{tx}(W_x + L_x)^{-1}C_{xt}\rvert$	$\Lambda(n - s, s, q - s)$
$*\Lambda_{C\lvert D} = \lvert W_x\rvert / \lvert W_x + L_x\rvert *\Lambda_D$	$\Lambda(n - 2s, p - s, q - s)$
$*T_D^2 = \mathrm{tr}\{C_{tx}(L_x + W_x)^{-1}C_{xt}[C_{tx}(L_x + W_x)^{-1}W_x(L_x + W_x)^{-1}C_{xt}]^{-1}\} - s$	$T_o^2(n - s, s, q - s)$
$*T_{C\lvert D}^2 = \mathrm{tr}\ W_x^{-1}L_x - \mathrm{tr}[C_{tx}(L_x + W_x)^{-1}C_{xt}]^{-1}C_{tx}W_x^{-1}C_{xt} + s$	$T_o^2(n - 2s, p - s, q - s)$
$*V_D = \mathrm{tr}[C_{tx}(L_x + W_x)^{-1}C_{xt}]^{-1}C_{tx}(L_x + W_x)^{-1}L_x(L_x + W_x)^{-1}C_{xt}$	$V(n - s, s, q - s)$
$*V_{C\lvert D} = \mathrm{tr}(W_x + L_x)^{-1}L_x - \mathrm{tr}\{C_{tx}(L_x + W_x)^{-1}W_x(L_x + W_x)^{-1}L_x(L_x$ $+ W_x)^{-1}C_{xt}[C_{tx}(L_x + W_x)^{-1}W_x[L_x + W_x)^{-1}C_{xt}]^{-1}\}$	$V(n - 2s, p - s, q - s)$
$*r_D^2 = \max_{r^2}\{r^2 : \ \lvert -(1 - r^2)I_s + [C_{tx}(L_x + W_x)^{-1}C_{xt}]^{-1}C_{tx}(L_x + W_x)^{-1}$ $\cdot W_x(W_x + L_x)^{-1}C_{xt}\rvert = 0\}$	$r_1^2(n - s, s, q - s)$
$*r_{C\lvert D}^2 = \max_{r^2}\{r^2 : \ \lvert -(1 - r^2)\{L_x + W_x - C_{xt}[C_{tx}(L_x + W_x)^{-1}C_{xt}]^{-1}C_{tx}\} + W_x\rvert = 0\}$	$r_1^2(n - 2s, p - s, q - s)$