R E P O R T   R E S U M E S

ED 017 973   24   EA 001 099

LIKELIHOOD RATIO TESTS OF HYPOTHESES ON MULTIVARIATE POPULATIONS, VOLUME II, TEST OF HYPOTHESIS--STATISTICAL MODELS FOR THE EVALUATION AND INTERPRETATION OF EDUCATIONAL CRITERIA. PART 4.

BY- SAW, J.G.
VIRGINIA POLYTECHNIC INST., BLACKSBURG

REPORT NUMBER CRP-1132-PT-4   PUB DATE   64
REPORT NUMBER BR-5-0224
EDRS PRICE   MF-$0.50   HC-$2.84   69P.

DESCR: TORS- *HYPOTHESIS TESTING, *STATISTICAL ANALYSIS, *MATHEMATICAL MODELS, *STATISTICS, CLASSIFICATION, TESTS OF SIGNIFICANCE, ANALYSIS OF VARIANCE, DATA ANALYSIS, PROBABILITY, PREDICTIVE MEASUREMENT, MATHEMATICAL APPLICATIONS, PREDICTIVE VALIDITY, *MEASUREMENT TECHNIQUES, BLACKSBURG, DOOLITTLE METHOD,

THIS PAPER DEALS WITH SOME TESTS OF HYPOTHESIS FREQUENTLY ENCOUNTERED IN THE ANALYSIS OF MULTIVARIATE DATA. THE TYPE OF HYPOTHESIS CONSIDERED IS THAT WHICH THE STATISTICIAN CAN ANSWER IN THE NEGATIVE OR AFFIRMATIVE. THE DOOLITTLE METHOD MAKES IT POSSIBLE TO EVALUATE THE DETERMINANT OF A MATRIX OF HIGH ORDER, TO SOLVE A MATRIX EQUATION, OR TO INVERT A MATRIX OF HIGH ORDER. HYPOTHESES WHICH, IN THE UNIVARIATE CASE, WOULD LEAD TO A USE OF THE STUDENT T-DISTRIBUTION ARE ALSO PRESENTED. THE ANALYSIS OF DATA COLLECTED FROM A "DESIGNED EXPERIMENT" IS DISCUSSED, INCLUDING THE LATIN SQUARE, RANDOMIZED BLOCK, CROSS-CLASSIFICATION, AND FACTORIAL DESIGNS. AN ANALYSIS OF REGRESSION IS PRESENTED ALONG WITH A TEST OF HYPOTHESIS ON THE DISPERSION MATRIX. (HW)

~~No~~ BR No
PA 24

STATISTICAL MODELS FOR THE EVALUATION AND INTERPRETATION OF EDUCATIONAL CRITERIA

Cooperative Research Project Number 1132

by

J. G. Saw

Virginia Polytechnic Institute
Blacksburg, Virginia
1964

PART 4. LIKELIHOOD RATIO TESTS OF HYPOTHESES ON MULTIVARIATE POPULATIONS

VOLUME II. TEST OF HYPOTHESIS

The presentation of these notes was supported through the Cooperative Research Program of the Office of Education, U. S. Department of Health, Education, and Welfare.

ACKNOWLEDGEMENTS

This volume was written with the financial support of the U. S Department of Health, Education, and Welfare under Cooperative Research Project Number 1132.

INDEX

PREFACE

In this short volume, we deal with some tests of hypothesis which are frequently encountered in the analysis of multivariate data. The type of hypothesis considered is that which can be answered in the negative or affirmative by the statistician (with certain calculable probabilities of being wrong).

It will be recognized that this type of hypothesis covers but a small part of the statisticians work in the multivariate field. In the simplest of all cases we may be presented with a sequence of vector observations and asked to judge whether or not these vector observations could have been drawn from a population with vector mean $\underline{\mu}_o$. If the statistician judges 'no', then almost inevitably he will be asked "if the time population mean vector is $\underline{\mu}$ which you (the statistician) say is not $\underline{\mu}_o$, in which of its elements is $\underline{\mu}$ different from $\underline{\mu}_o$?" The best the statistician can do in answer to this question is to make an educated guess. For suppose the vector has m elements and $\underline{\mu} - \underline{\mu}_o = \underline{e}$. The statistician has judged that the vector $\underline{e}$ is not the vector of all zeros and by controlling his first kind of error to the usual 5%, he will be wrong on the average one time in twenty situations where $\underline{\mu}$ really is $\underline{\mu}_o$. But in answer to the follow-up question he is being asked to pick the correct alternative of 2^m-1 possible alternatives:

that $\underline{e}$ differs from $\underline{0}$ only in its i-th position for some i

that $\underline{e}$ differs from $\underline{0}$ only in its i-th and j-th position for some i and j ($i \neq j$)

----

etcetera

----

that $\underline{e}$ differs from $\underline{0}$ in all but its i-th position for some i. There are no statistical techniques which enable the statistician to announce (for example) "$\underline{\mu}$ differs from $\underline{\mu}_o$ in its first and third position only and there's a 5% chance that I'm wrong." The heart of the matter lies in the theoretical impossibility of putting a probability of error on the statement. In connexion with this, the reader should consult section 11.6; page 258.

Nevertheless, certain proceedures are available which can be described only as logical proceedures upon which to base an educated guess. It is these proceedures which are not developed in this volume though the reader is referred to Part III of this contract written by Rolf Bargmann for a detailed discussion, illustration and analysis of a similar type of multi-alternative problem and to Chapter 15 of these notes for an introduction to that problem.

CHAPTER 10
THE DOOLITTLE METHOD

10.1 Introduction

In the process of the analysis of actual experimental data, it will frequently be necessary to evaluate the determinant of a matrix of high order; to solve a matrix equation or to invert a matrix of high order. The Doolittle method is well established as a computational technique which lends itself to the use of a desk calculator or electronic computor and is widely used in the analysis of multivariate data. The evaluation of a determinant, matrix invertion, and solution of the matrix equation is accomplished through a series of systematic elementary mathematical operations.

10.2 The forward Doolittle process.

Given a symmetric, positive definite matrix A with known elements, we may wish to

(a) evaluate $|A|$

(b) determine A^{-1}

(c) determine Φ satisfying $A\Phi = B$ (B given) .

We shall hold that A be mxm; in problem (c) above, B may be mxk ($k \leq m$) so that Φ is also mxk. It is noted that problems (b) and (c) are not dissimilar for if we set B=I, then $\Phi = A^{-1}$. We will discuss problem (c) and give the methods for (a) and (b) as special cases.

Set out algebraically, the computation table has the appearance:

a_{11}	a_{12}	a_{13}	a_{14}	....	a_{1m}	b_{11}	b_{12}	b_{13}	....	b_{1k}	x_1
a^{*}_{11}	a^{*}_{12}	a^{*}_{13}	a^{*}_{14}	....	a^{*}_{1m}	b^{*}_{11}	b^{*}_{12}	b^{*}_{13}	....	b^{*}_{1k}	x^{*}_1
1	a^{**}_{12}	a^{**}_{13}	a^{**}_{14}	....	a^{**}_{1m}	b^{**}_{11}	b^{**}_{12}	b^{**}_{13}	....	b^{**}_{1k}	x^{**}_1
(a_{21})	a_{22}	a_{23}	a_{24}	....	a_{2m}	b_{21}	b_{22}	b_{23}	....	b_{2k}	x_2
(0)	a^{*}_{22}	a^{*}_{23}	a^{*}_{24}	....	a^{*}_{2m}	b^{*}_{21}	b^{*}_{22}	b^{*}_{23}	....	b^{*}_{2k}	x^{*}_2
(0)	1	a^{**}_{23}	a^{**}_{24}	....	a^{**}_{2m}	b^{**}_{21}	b^{**}_{22}	b^{**}_{23}	....	b^{**}_{2k}	x^{**}_2
(a_{31})	(a_{32})	a_{33}	a_{34}	....	a_{3m}	b_{31}	b_{32}	b_{33}	....	b_{3k}	x_3
(0)	(0)	a^{*}_{33}	a^{*}_{34}	....	a^{*}_{3m}	b^{*}_{31}	b^{*}_{32}	b^{*}_{33}	....	b^{*}_{3k}	x^{*}_3
(0)	(0)	1	a^{**}_{34}	....	a^{**}_{3m}	b^{**}_{31}	b^{**}_{32}	b^{**}_{33}	....	b^{**}_{3k}	x^{**}_3
								etcetera			
					a_{mm}	b_{m1}	b_{m2}	b_{m3}	....	b_{mk}	x_m
					a^{*}_{mm}	b^{*}_{m1}	b^{*}_{m2}	b^{*}_{m3}	....	b^{*}_{mk}	x^{*}_m
					a^{**}_{mm}	b^{**}_{m1}	b^{**}_{m2}	b^{*}_{m3}	....	b^{**}_{mk}	x^{**}_m

It is noted that the first row of the table is the first row of A, followed by the first row of B, followed by a "check" entry x_1 which is actually the sum of all elements (m+k of them) in the first rows of A and B, so that

(10.2.1) $$x_1 = \sum_{j=1}^{m} a_{1j} + \sum_{j=1}^{k} b_{1j}$$

the second row of the table is the first row repeated [in practice

the first row could be omitted; it is included here for symmetry].
This second row which is the first starred row is defined then by:

$$(10.2.2)\qquad \begin{array}{ll} a^*_{1j} = a_{1j} & j = 1, 2, \ldots, m \\ b^*_{1j} = b_{1j} & j = 1, 2, \ldots, k \\ x^*_1 = x_1 & \end{array}$$

The third row (first double-starred row) is produced by dividing all elements in the proceeding row by the leading element of that row (that is, by a^*_{11}). The first row of double-starred elements are defined then by:

$$(10.2.3)\qquad \begin{array}{ll} a^*_{1j} = a^*_{1j}/a^*_{11} & j = 1, 2, \ldots, m \\ b^{**}_{1j} = b^*_{1j}/a^*_{11} & j = 1, 2, \ldots, k \\ x^{**}_1 = x^*_1/a^*_{11} & \end{array}$$

The fourth row is produced by writing down the elements of the second row of A followed by the elments of the second row of B, followed by the "check" entry x_2, where

$$(10.2.4)\qquad x_2 = \sum_{j=1}^{m} a_{2j} + \sum_{j=1}^{k} b_{2j} \quad .$$

In practice the element a_{21} is not written in (it appears in brackets in the table since the elements which will appear under it are zero and do not really enter into the computations). The second starred row is computed via the equations:-

$$a_{2j}^{*} = a_{2j} - a_{12}^{*}a_{1j}^{*} \qquad j = 2, 3, \ldots, m$$
(10.2.5) $$b_{2j}^{*} = b_{2j} - a_{12}^{*}b_{1j}^{**} \qquad j = 1, 2, \ldots, k$$
$$x_2^{*} = x_2 - a_{12}^{*}x_1$$

It is observed that $a_{21}^{*} = a_{21} - a_{12}^{*}a_{11}^{*} = a_{21} - a_{12} \equiv 0$ by the symmetry of A. There is no point in entering this zero in the table in practice.

The second row of double-starred elements is obtained by dividing the preceeding row by the element a_{22}^{*}, thus

$$a_{2j}^{**} = a_{2j}^{*}/a_{22}^{*} \qquad j = 2, 3, \ldots, m$$
(10.2.6) $$b_{2j}^{**} = b_{2j}^{*}/a_{22}^{*} \qquad j = 1, 2, \ldots, k$$
$$x_2^{**} = x_2^{*}/a_{22}^{*}$$

The seventh row is the third row of A, followed by the third row of B, followed by "check" entry x_3 defined by

(10.2.7) $$x_3 = \sum_{j=1}^{m} a_{3j} + \sum_{j=1}^{k} b_{3j}$$

The third starred row is obtained via the equations

$$a_{3j}^{*} = a_{3j} - a_{23}^{*}a_{2j}^{**} - a_{13}^{*}a_{1j}^{**} \qquad j = 3, 4, \ldots, m$$
(10.2.8) $$b_{3j}^{*} = b_{3j} - a_{23}^{*}b_{2j}^{**} - a_{13}^{*}b_{1j}^{**} \qquad j = 1, 2, \ldots, k$$
$$x_3^{*} = x_3 - a_{23}^{*}x_2^{**} - a_{13}^{*}x_1^{**}$$

It can be shown that $a_{31}^{*} = a_{32}^{*} \equiv 0$, so that in practice a_{31} and a_{32} and any entry below these elements are omitted from the table. Notice that a_{23}^{*} and a_{13}^{*} act as multipliers for all elements in in the third block, third starred row so that it is a good idea to pencil a circle round these two (a^*_{23} and a^*_{13}) when working in the third block.

The third row of double starred elements are obtained by dividing the preceeding row by a^*_{33} so that

$$(10.2.9)\qquad \begin{aligned} a^{**}_{3j} &= a^*_{3j}/a^*_{33} & j &= 3, 4, \ldots, m \\ b^{**}_{3j} &= b^*_{3j}/a^*_{33} & j &= 1, 2, \ldots, k \\ x^{**}_3 &= x^*_3/a^*_{33} \end{aligned}$$

In general the first row of the r-th block ($\overline{3r-2}$-th row) is the r-th row of A, followed by the r-th row of B, followed by the "check" entry x_r given by

$$x_r = \sum_{j=1}^{m} a_{rj} + \sum_{j=1}^{k} b_{rj}$$

and in practice the elements $a_{r1}, a_{r2}, \ldots, a_{rr-1}$ are not entered since elements $\{a^*_{ri}\}_{i=1}^{i=r-1}$ turn out to be zero. The r-th starred entries are given by

$$(10.2.10)\qquad \begin{aligned} a^*_{rj} &= a_{rj} - \sum_{i=1}^{r-1} a^*_{ir}a^{**}_{ij} & j &= r, r+1, \ldots, m \\ b^*_{rj} &= b_{rj} - \sum_{i=1}^{r-1} a^*_{ir}b^{**}_{ij} & j &= 1, 2, \ldots, k \\ x^*_r &= x_r - \sum_{i=1}^{r-1} a^*_{ir}x_i \end{aligned}$$

again; when working in the r-th block it is a good idea to pencil a circle round elements $a^*_{1r}, a^*_{2r}, \ldots, a^*_{r-1:r}$ since they occur in all the multiplications. The r-th double starred row is given by

$$(10.2.11)\qquad \begin{aligned} a^{**}_{rj} &= a^{*}_{rj}/a^{*}_{rr} && j = r, r+1, \ldots, m \\ b^{**}_{rj} &= b^{*}_{rj}/a^{*}_{rr} && j = 1, 2, \ldots, k \\ x^{**}_{r} &= x^{*}_{r}/a^{*}_{rr} \end{aligned}$$

The process is repeated until all rows of A are exhausted (producing then m blocks of three line entries).

As a check on the computations, it is noted that

$$(10.2.12)\qquad \begin{aligned} x^{**}_{r} &= \sum_{j=1}^{m} a^{**}_{rj} + \sum_{j=1}^{k} b^{**}_{rj} \\ &= \sum_{j=r}^{m} a^{**}_{rj} + \sum_{j=1}^{k} b^{**}_{rj} \quad (\text{since } \{a^{**}_{rj} = 0\}_{j=1}^{j=r-1}) \end{aligned}$$

This completes the so-called "forward" part of the Doolittle process and it is instructive to consider what has been done algebraically. It is easy to verify that in order to go from $A = (a_{ij})$ to $A^{*} = (a^{*}_{ij})$, we have premultiplied A by a lower triangular matrix with ones on the diagonal. Calling this matrix F, we have

$$(10.2.13)\qquad F = \begin{matrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ f_{21} & 1 & 0 & 0 & \cdots & 0 \\ f_{31} & f_{32} & 1 & 0 & \cdots & 0 \\ f_{41} & f_{42} & f_{43} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ f_{m1} & f_{m2} & f_{m3} & f_{m4} & \cdots & 1 \end{matrix}$$

In fact, F is the product of matrices

$$F = J_m \, J_{m-1} \, J_{m-2} \; \cdots\cdots \; J_3 \, J_2 \, J_1 \tag{10.2.14}$$

where

$$J_1 = I$$

$$J_2 = \begin{matrix} 1 & 0 & 0 & 0 & \cdots\cdots & 0 \\ \frac{-a^*_{12}}{a^*_{11}} & 1 & 0 & 0 & \cdots\cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots\cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots\cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & & 1 \end{matrix}$$

and in general J_r is the identity matrix with the r-th row replaced by

$$\frac{-a^*_{1r}}{a^*_{11}}, \; \frac{-a^*_{2r}}{a^*_{22}}, \; \frac{-a^*_{3r}}{a^*_{33}}, \; \ldots\ldots, \; \frac{-a^*_{r-1:r}}{a^*_{r-1:r-1}}, \; 1, \; 000..0$$

Now exactly the same operations are carried out on B as those on A; accordingly if we define $B^* = (b^*_{ij})$, then

$$\left\{ \begin{matrix} FA = A^* \\ \\ FB = B^* \end{matrix} \right. \qquad \text{or } F(A \vdots B) = (A^* \vdots B^*) \tag{10.2.15}$$

To produce the double-starred elements, we divide the r-th row of $(A \vdots B)$ by a^*_{rr}. Define the diagonal matrix whose r-th diagonal element is a^*_{rr} by D

then

$$\text{(10.2.16)} \qquad \begin{aligned} D^{-1}A^* &= D^{-1}FA = A^{**} \\ D^{-1}B^* &= D^{-1}FB = B^{**} \end{aligned}$$

10.3 The solution of $A\Phi=B$.

We are now in a position to determine Φ by simple arithmetic operations. Since

$$\text{(10.3.1)} \qquad A\Phi = B,$$

then if F be the forward Doolittle process on A carrying B, then

$$\text{(10.3.2)} \qquad D^{-1}FA\Phi = D^{-1}FB$$

or

$$\text{(10.3.3)} \qquad A^{**}\Phi = B^{**}$$

The solution for Φ is simple by virtue of the fact that A** is upper triangular. Writing equation (10.3.3) out in more detail:

$$\text{(10.3.4)} \quad \begin{bmatrix} 1 & a^{**}_{12} & a^{**}_{13} & \cdots & a^{**}_{1m-1} & a^{**}_{1m} \\ 0 & 1 & a^{**}_{23} & \cdots & a^{**}_{2m-1} & a^{**}_{2m} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & a^{**}_{m-1:m} \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \begin{bmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1k} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2k} \\ \vdots & \vdots & & \vdots \\ \phi_{m-1:1} & \phi_{m-1:2} & \cdots & \phi_{m-1:k} \\ \phi_{m1} & \phi_{m2} & \cdots & \phi_{mk} \end{bmatrix}$$

$$= \begin{bmatrix} b^{**}_{11} & b^{**}_{12} & \cdots & b^{**}_{1k} \\ b^{**}_{21} & b^{**}_{22} & \cdots & b^{**}_{2k} \\ \vdots & \vdots & & \vdots \\ b^{**}_{m-1:1} & b^{**}_{m-1:2} & \cdots & b^{**}_{m-1:k} \\ b^{**}_{m1} & b^{**}_{m2} & \cdots & b^{**}_{mk} \end{bmatrix}$$

Multiplying the last row of A** into Φ we have

(10.3.5) $$(\phi_{m1}\ \phi_{m2}\ \cdots\ \phi_{mk}) = (b^{**}_{m1}\ b^{**}_{m2}\ \cdots\ b^{**}_{mk}) .$$

Multiplying the penultimate row of A** into Φ we have

(10.3.6) $$(\phi_{m-1:1}\ \phi_{m-1:2}\ \cdots\ \phi_{m-1:k}) + a^{**}_{m-1:m}(\phi_{m:1}\ \phi_{m:2}\cdots\ \phi_{m-1:k})$$
$$= (b^{**}_{m-1:1}\ b^{**}_{m-1:2}\ \cdots\ b^{**}_{m-1:k})$$

so that $\{\phi_{m-1:j}\}_1^k$ are quickly determined.

Similarly for all rows of Φ ending with the first row of Φ. An example is given in section 10.5.

10.4 The determinant and inverse of A.

Returning to equation (10.2.15), we have, after taking determinants of left- and right-hand sides,

(10.4.1) $$|F||A| = |A^*|$$

but $|F| = 1$ and $|A^*| = a^*_{11}\ a^*_{22}\ \cdots\ a^*_{mm}$, therefore

(10.4.2) $$|A| = a^*_{11}\ a^*_{22}\ \cdots\ a^*_{mm} .$$

It is important also to observe that if FA = A*, then

$$\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ f_{21} & 1 & 0 & \cdots & 0 \\ f_{31} & f_{32} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ f_{r1} & f_{r2} & f_{r3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1r} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2r} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3r} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{r1} & a_{r2} & a_{r3} & \cdots & a_{rr} \end{bmatrix} = \begin{bmatrix} a^*_{11} & a^*_{22} & a^*_{23} & \cdots & a^*_{1r} \\ 0 & a^*_{22} & a^*_{23} & \cdots & a^*_{2r} \\ 0 & 0 & a^*_{33} & \cdots & a^*_{3r} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & a_{rr} \end{bmatrix}$$

so that

(10.4.3)
$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1r} \\ a_{21} & a_{22} & \cdots & a_{2r} \\ \vdots & \vdots & & \vdots \\ a_{r1} & a_{r2} & \cdots & a_{rr} \end{vmatrix} = a^*_{11}\, a^*_{22} \cdots a^*_{rr} \quad .$$

This immediately yields another important result. Let A be partitioned:

(10.4.4)
$$A = \left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right]$$

with A_{11} and A_{22} $\overline{pxp}$ and $\overline{qxq}$ respectively, where $p+q = m$. Then (by clockwise rule)

(10.4.5)
$$|A| = |A_{11}|\,|A_{22}-A_{21}A_{11}^{-1}A_{12}| \quad .$$

Using (10.4.2) and (10.4.3)

(10.4.6)
$$|A_{22}-A_{21}A_{11}^{-1}A_{12}| = a^*_{r+1:r+1}\; a^*_{r+2:r+2} \cdots a^*_{mm} \quad .$$

Turning now to the inverse of A, we carry I (mxm) rather than general B. In our notation, we have:

(10.4.7)
$$\begin{cases} F(A \vdots I) = (A^* \vdots I^*) \\ D^{-1}(A^* \vdots I^*) = (A^{**} \vdots I^{**}) \end{cases} \quad .$$

Notice that

(10.4.8)
$$\begin{cases} I^* = F \\ I^{**} = D^{-1}F \end{cases} \quad .$$

Now, postmultiplying the first of equations (10.2.15) by F', we have

$$(10.4.9) \qquad FAF' = A^*F' \quad .$$

Since A^* and F' are both upper triangular, then so also is their product; but FAF' ($=A^*F'$) is symmetric. Since A^*F' is upper triangular and symmetric, it must be diagonal. Clearly $(A^*F')_{rr} = a^*_{rr}$, therefore

$$(10.4.10) \qquad FAF' = A^*F' = D \quad ,$$

so that

$$(10.4.11) \qquad \begin{aligned} A^{-1} &= F'D^{-1}F \\ &= (I^*)'(I^{**}) \quad , \end{aligned}$$

so that

$$(10.4.12) \qquad (A^{-1})_{\alpha\beta} = ((I^*)_{.\alpha})'((I^{**})_{.\beta});$$

that is the (α,β)-th element of A^{-1} is the inner product of the α-th column of I^* with the β-th column of I^{**}. An example is given in the next section.

10.5 Examples of the problems discussed in section 1.

Problem 1

Let it be required to solve for Φ from the equation $A\Phi = B$ when

$$A = \begin{bmatrix} 361 & 532 & 703 & 475 & 0 \\ 532 & 1145 & 1036 & 1175 & 0 \\ 703 & 1036 & 1469 & 1025 & 100 \\ 475 & 1175 & 1025 & 1975 & 725 \\ 0 & 0 & 100 & 725 & 2094 \end{bmatrix}$$

and

$$B = \begin{bmatrix} 3420 & 2489 \\ 5990 & 5093 \\ 7360 & 5447 \\ 8950 & 8875 \\ 5938 & 6463 \end{bmatrix}$$

A					B		X
361. 361. 1.	532. 532. 1.47368421	703. 703. 1.94736842	475. 475. 1.31578947	0. 0. 0.	3420. 3420. 9.47368421	2489. 2489. 6.89473684	7980. 7980. 22.10526316
	1145. 361.000000 1.	1036. 0. 0.	1175. 475.000002 1.31578947	0. 0. 0.	5990. 950.000000 2.63157895	5093. 1425.000001 3.94736842	14971. 3210.999999 8.89473684
		1469. 100.000000 1.00000000	1025. 100.000000 1.00000000	100. 100.000000 1.00000000	7360. 700.000000 7.00000000	5447. 600.000000 6.00000000	17140. 1600.000000 16.00000000
			1975. 625.000001 1.00000000	725. 625.000001 1.00000000	8950. 2500.000000 4.00000000	8875. 3125.000000 5.00000000	23200. 6875. 11.00000000
				2094. 1369.000000 1.00000000	5938. 2738.000000 2.00000000	6463. 2738.000000 2.00000000	15320. 6845.000000 5.00000000

The figures are held correct to the number of decimal places shown.

Having performed the forward Doolittle process, the equation $A\Phi = B$ becomes $A^{**}\Phi = B^{**}$ where

$$A^{**} = \begin{bmatrix} 1.00000000 & 1.47368421 & 1.94736842 & 1.31578947 & 0.00000000 \\ \text{zero} & 1.00000000 & 0.00000000 & 1.31578947 & 0.00000000 \\ \text{zero} & \text{zero} & 1.00000000 & 1.00000000 & 1.00000000 \\ \text{zero} & \text{zero} & \text{zero} & 1.00000000 & 1.00000000 \\ \text{zero} & \text{zero} & \text{zero} & \text{zero} & 1.00000000 \end{bmatrix}$$

$$B^{**} = \begin{bmatrix} 9.47368421 & 6.89473684 \\ 2.63157895 & 3.94736842 \\ 7.00000000 & 6.00000000 \\ 4.00000000 & 5.00000000 \\ 2.00000000 & 2.00000000 \end{bmatrix}$$

We have immediately

$$\phi_{51} = 2.00000000$$

$$\phi_{52} = 2.00000000$$

$$\phi_{41} = 4.00000000-(1.00000000)(2.00000000) = 2.00000000$$

$$\phi_{42} = 5.00000000-(1.00000000)(2.00000000) = 3.00000000$$

$$\phi_{31} = 7.00000000-(1.00000000)(2.00000000) -(1.00000000)(2.00000000) = 3.00000000$$

$$\phi_{32} = 6.00000000-(1.00000000)(3.00000000) -(1.00000000)(2.00000000) = 1.00000000$$

and similarly

$\phi_{21} = 0.00000000$

$\phi_{22} = 0.00000000$

$\phi_{11} = 1.00000000$

$\phi_{12} = 1.00000000$

Problem II

It is required to evaluate the determinant of matrix A when A is

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 5 & 4 & 4 & 10 \\ 3 & 4 & 15 & 24 & 17 \\ 4 & 4 & 24 & 45 & 44 \\ 5 & 10 & 17 & 44 & 110 \end{bmatrix}$$

(this matrix has been chosen so that the forward Doolittle can be performed without the need of a desk calculator).

The computations proceed:

		A			X
1	2	3	4	5	15
1	2	3	4	5	15
1	2	3	4	5	15
	5	4	4	10	25
	1	-2	-4	0	-5
	1	-2	-4	0	-5
		15	24	17	63
		2	4	2	8
		1	2	1	4
			45	44	121
			5	20	25
			1	4	5
				110	186
				3	3
				1	1

and

$$A = \prod_{i=1}^{5} a^*_{ii} = 1\times1\times2\times5\times3 = 30 \quad .$$

Problem III

It is required to obtain the inverse of A, symmetric, when

$$A = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 5 & -1 & 2 & 3 \\ 1 & -1 & 3 & 0 & 0 \\ 0 & 2 & 0 & 3 & 1 \\ 0 & 3 & 0 & 1 & 10 \end{bmatrix}$$

(Here again, for demonstration purposes, A has been chosen so that the Doolittle proceedure does not require the use of a desk calculator).

The required table develops as follows:

←		A		→	←		I		→	X
1	1	1	0	0	1	0	0	0	0	4
1	1	1	0	0	1	0	0	0	0	4
1	1	1	0	0	1	0	0	0	0	4
	5	-1	2	3	0	1	0	0	0	11
	4	-2	2	3	-1	1	0	0	0	7
	1	-1/2	1/2	3/4	-1/4	1/4	0	0	0	7/4
		3	0	0	0	0	1	0	0	4
		1	1	3/2	-3/2	1/2	1	0	0	7/2
		1	1	3/2	-3/2	1/2	1	0	0	7/2
			3	1	0	0	0	1	0	7
			1	-2	2	-1	-1	1	0	0
			1	-2	2	-1	-1	1	0	0
				10	0	0	0	0	1	15
				3/2	7	-7/2	-7/2	2	1	9/2
				1	14/3	-7/3	-7/3	4/3	2/3	3

and the values of $(A^{-1})_{\alpha\beta}$ are obtained (equation 10.4.17) directly from the entries under "I"; thus

$$(A^{-1})_{11} = (1)(1) + (-1)(-\tfrac{1}{4}) + (-\tfrac{3}{2})(-\tfrac{3}{2}) + (2)(2) + (7)(\tfrac{14}{3}) = \frac{482}{12}$$

$$(A^{-1})_{12} = (1)(0) + (-1)(\tfrac{1}{4}) + (-\tfrac{3}{2})(\tfrac{1}{2}) + (2)(-1) + (7)(-\tfrac{7}{3}) = -\frac{232}{12}$$

etcetera

$$A^{-1} = \frac{1}{6}\begin{bmatrix} 241 & -116 & -119 & 68 & 28 \\ -116 & 58 & 58 & -34 & -14 \\ -119 & 58 & 61 & -34 & -14 \\ 68 & -34 & -34 & 22 & 8 \\ 28 & -14 & -14 & 8 & 4 \end{bmatrix} .$$

It is advisable as a final check to perform the product AA^{-1} to make sure the identity matrix does indeed result.

10.6 The computation of $B'A^{-1}B$. (A and B given).

Perform the forward Doolittle on A carrying B:

$$(10.6.1) \qquad \begin{aligned} F(A \vdots B) &= (A^* \vdots B^*) \\ D^{-1}(A^* \vdots B^*) &= (A^{**} \vdots B^{**}) \end{aligned} .$$

Now since $A^{-1} = F'D^{-1}F$ (equation 10.4.16)

$$(10.6.2) \qquad \begin{aligned} B'A^{-1}B &= B'F'D^{-1}FB \\ &= (FB)'D^{-1}FB = (B^*)'(B^{**}) \end{aligned}$$

so that

$$(10.6.4) \qquad (B'A^{-1}B)_{\alpha\beta} = ((B^*)_{.\alpha})'((B^{**})_{.\beta})$$

that is, the (α,β)-th element of $B'A^{-1}B$ is the inner product of the α-th column of B^* with the β-th column of B^{**} .

Problem IV

Obtain the value of the quadratic form $\underline{x}'A^{-1}\underline{x}$ when

$$A=\begin{bmatrix} 1 & 2 & 3 & 3 & 3 \\ 2 & 9 & 6 & 9 & 0 \\ 3 & 6 & 14 & 10 & 10 \\ 3 & 9 & 10 & 12 & 12 \\ 3 & 0 & 10 & 12 & 18 \end{bmatrix}$$

$\underline{x}' = (1 \quad 0 \quad 4 \quad 2 \quad 3)$.

The required table develops as follows:

A					$\underline{x}$	X
1	2	3	3	3	1	13
1	2	3	3	3	1	13
1	2	3	3	3	1	13
	9	6	9	0	0	26
	5	0	3	-6	-2	0
	1	0	0.6	-1.2	-0.4	0
		14	10	10	4	47
		5	1	1	1	8
		1	0.2	0.2	0.2	1.6
			12	12	2	48
			1	6.4	0	7.4
			1	6.4	0	7.4
				18	3	46
				-39.36	-2.6	-41.96
				1	$\frac{260}{3936}$	$\frac{4196}{3936}$

The value of $\underline{x}'A^{-1}\underline{x}$ is, from the entries under $\underline{x}$, (1)(1) + (-2)(-0.4) + (1)(0.2) + (0)(0) = $(2.6)\frac{(260)}{(3936)} = \frac{7196}{3936}$ = 1.828252 (=$\underline{x}'A^{-1}\underline{x}$).

CHAPTER 11

TESTS OF SIMPLE LINEAR HYPOTHESES

11.1 Introduction

In this chapter we deal with hypotheses which in the univariate case would lead to a use of the student t-distribution. In its most general statement we have, say, g groups or populations and n_i vector observations from the i-th group. Our observations will be designated $\underline{x}_{ij}$ ($j=1,2,\ldots,n_i$; $i=1,2,\ldots,g$) and it will be assumed

$$\underline{x}_{ij} \frown N_m(\underline{\mu}_i;V) \tag{11.1.1}$$

the $N = \sum_{i=1}^{g} n_i$ vectors $\{\underline{x}_{ij}\}$ being mutually independent.

The most general simple hypothesis is:

$$H_o: \quad \beta_1\underline{\mu}_1+\beta_2\underline{\mu}_2+\ldots+\beta_g\underline{\mu}_g = \underline{\mu}_o \tag{11.1.2}$$

where $\{\beta_i\}_1^g$ are a set of specified scalar constants and where $\underline{\mu}_o$ is a specified mx1 vector of constants (often the vector of zeros).

The hypothesis expressed in (11.1.2) includes the following more familiar cases:

(a) H_o: (one group) $\underline{\mu} = \underline{\mu}_o$ $(g=1;\beta_1=1)$

(b) H_o: (two groups) $\underline{\mu}_1 = \underline{\mu}_2$ $(g=2;\beta_1=1;\beta_2=-1;\underline{\mu}_o=\underline{0})$

and actually includes the regression model but this will be treated separately.

11.2 The test statistic for the general simple linear hypothesis.

Given the observations described in the introduction, that is, given

$$(11.2.1) \qquad \underline{x}_{ij} \frown N(\underline{\mu}_i; V) \qquad \begin{array}{l} j = 1, 2, \ldots, n_i \\ i = 1, 2, \ldots, g \end{array} ,$$

it is required to test

$$(11.2.2) \qquad H_o: \sum_{i=1}^{g} \beta_i \underline{\mu}_i = \underline{\mu}_o$$

for specified $\{\beta_i\}_1^g$ and $\underline{\mu}_o$.

Method:

Obtain the sample product—cross product matrix for the i-th group

$$(11.2.3) \qquad C_i = \sum_{j=1}^{n_i} (\underline{x}_{ij} - \overline{\underline{x}}_{i.})(\underline{x}_{ij} - \overline{\underline{x}}_{i.})'$$

$$= \sum_{j=1}^{n_i} \underline{x}_{ij}\underline{x}'_{ij} - n_i \overline{\underline{x}}_{i.} \overline{\underline{x}}'_{i.}$$

where

$$\overline{\underline{x}}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} \underline{x}_{ij}$$

and construct

$$(11.2.4) \qquad C = C_1 + C_2 + \ldots + C_g$$

so that C is an unbiassed estimate of (N-g)V.

Clearly

$$(11.2.5) \qquad \underline{d} = \left(\sum_{i=1}^{n_i} \beta_i \overline{\underline{x}}_{i.} \right) - \underline{\mu}_o$$

which has expectation $\left(\sum_{i=1}^{n_i} \beta_i \underline{\bar{x}}_i\right) - \underline{\mu}_o$ is a measure of departure from H_o having $\underline{0}$ expectation if H_o be true. The statistic $\underline{d}$ is actually a vector randum variable having the distribution

(11.2.6) $$\underline{d} \frown N_m(\sum \beta_i \underline{\mu}_i - \underline{\mu}_o;\ (\sum \frac{(\beta_i)^2}{n_i})\ V)$$

so that "standardization to V" is effected by dividing $\underline{d}$ by $\sqrt{\sum(\beta_i^2 / n_i)}$. <u>The test criterion</u> becomes

(11.2.7) $$\mathcal{L} = (\sum(\beta_i^2 / n_i))^{-1}\ \underline{d}' C^{-1} \underline{d}$$

and <u>the critical region</u> of size α is given by

(11.2.8) $$\mathcal{L} > \frac{m}{N-g+1-m} F^{(\alpha)}_{m:N-g+1-m}$$

where $F^{(\alpha)}_{m:N-g+1-m}$ is the point of the $F_{m:N-g+1-m}$ density which cuts off 100 α% above. $\mathcal{L}$ is computed using the Doolittle proceedure (see Section 10).

<u>11.3 Simple test on the mean of a single population</u>

In the development of Section 11.2, we replace g by 1, set $\beta_1 = 1$, and observe

(11.3.1) $$\underline{x}_j \frown N_m(\underline{\mu};V) \qquad j = 1, \ldots, n.$$

It is required to test H_o: $\underline{\mu} = \underline{\mu}_o$. <u>The test criterion</u> becomes

(11.3.2) $$\mathcal{L} = n(\underline{\bar{x}} - \underline{\mu}_o)'\ C^{-1}\ (\underline{\bar{x}} - \underline{\mu}_o)$$

where

$$(11.3.3) \qquad C = \sum_{j=1}^{n} (\underline{x}_j - \underline{\bar{x}})(\underline{x}_j - \underline{\bar{x}})'$$

$$= \sum_{j=1}^{n} \underline{x}_j \underline{x}_j' - n\underline{\bar{x}}\underline{\bar{x}}'$$

$\mathcal{L}$ is computed using the Doolittle proceedure (see Section 10).

The critical region of size α is given by

$$(11.3.4) \qquad \mathcal{L} > \frac{m}{n-m} F^{(\alpha)}_{m:n-m}$$

11.4 A comparison of the means of two populations.

In the development of Section 11.2, we replace g by 2 and set $\beta_1 = 1$; $\beta_2 = -1$; $\underline{\mu}_o = \underline{0}$ and observe

$$(11.4.1) \qquad \begin{cases} \underline{x}_{1j} \frown N_m(\underline{\mu}_1; V) & j = 1, \ldots, n_1 \\ \underline{x}_{2j} \frown N_m(\underline{\mu}_2; V) & j = 1, \ldots, n_2 \end{cases} .$$

It is required to test H_o: $\underline{\mu}_1 = \underline{\mu}_2$ (i.e., $\underline{\mu}_1 - \underline{\mu}_2 = \underline{0}$). The standardizing factor is $\frac{1}{n_1} + \frac{1}{n_2} = \frac{n_1+n_2}{n_1 n_2}$ so that the test criterion becomes

$$(11.4.2) \qquad \mathcal{L} = \frac{n_1+n_2}{n_1 n_2} (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' C^{-1} (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)$$

where

$$(11.4.3) \qquad C = \sum_{j=1}^{n_1}(\underline{x}_{1j}-\underline{\bar{x}}_{1.})(\underline{x}_{1j}-\underline{\bar{x}}_{1.})' + \sum_{j=1}^{n_2}(\underline{x}_{2j}-\underline{\bar{x}}_{2.})(\underline{x}_{2j}-\underline{\bar{x}}_{2.})'$$

$$= \sum_{j=1}^{n_1}\underline{x}_{1j}\underline{x}'_{1j} + \sum_{j=1}^{n_2}\underline{x}_{2j}\underline{x}'_{2j} - n_1\underline{\bar{x}}_{1.}\underline{\bar{x}}'_{1.} - n_2\underline{\bar{x}}_{2.}\underline{\bar{x}}'_{2.}$$

with, of course, $\underline{\bar{x}}_{1.} = \sum_{j=1}^{n_1}\underline{x}_{1j} \Big/ n_1$; $\underline{\bar{x}}_{2.} = \sum_{j=1}^{n_2}\underline{x}_{2j} \Big/ n_2$.

$\mathcal{L}$ is computed using the Doolittle proceedures (see Section 10).

The critical region of size α is given by

$$(11.4.4) \qquad \mathcal{L} > \frac{m}{n_1+n_2-1-m} F^{(\alpha)}_{m:n_1+n_2-1-m} \quad .$$

11.5 An example: Testing a relationship between elements within the mean vector.

A task is performed on each of five successive days by 12 individuals (these individuals forming a fairly homogeneous group). It is noticeable that the "time to complete the task" decreases with each day due, most likely, to the experience gained. It is required to test the hypothesis that the time to complete the task is decreasing at a constant rate for each individual (the rate almost certainly differs for different individuals). The data are recorded below:

Time To Complete Task (Minutes)

Individual	Day 1	Day 2	Day 3	Day 4	Day 5
1	10.6	9.7	8.4	7.4	6.3
2	8.3	8.1	8.2	8.3	7.6
3	8.5	7.7	7.5	6.9	6.4
4	9.0	8.9	8.7	8.1	7.6
5	13.3	12.6	11.5	10.5	9.3
6	17.0	14.8	12.3	10.3	7.8
7	8.0	8.2	7.8	7.4	7.4
8	12.1	12.0	11.4	10.6	10.5
9	15.0	13.8	12.7	11.9	10.6
10	13.8	12.1	10.9	9.0	7.0
11	13.7	13.3	12.4	11.7	10.9
12	11.3	10.6	9.1	8.0	6.7

The data collected is say $\underline{y}_t$, t = 1, 2, ..., 12, with $\underline{y}_t$ the five entry vector for individual t. The mean vector $\underline{\mu}_t = \mathcal{E}(\underline{y}_t)$ should, under H_o, satisfy

$$H_o: \mu_{1t}-\mu_{2t} = \mu_{2t}-\mu_{3t} = \mu_{3t}-\mu_{4t} = \mu_{4t}-\mu_{5t} \qquad \text{(all t)}$$

or, alternatively

$$H_o: \left.\begin{array}{l} \mu_{1t} - 2\mu_{2t} + \mu_{2t} = 0 \\ \mu_{2t} - 2\mu_{3t} + \mu_{4t} = 0 \\ \mu_{3t} - 2\mu_{4t} + \mu_{5t} = 0 \end{array}\right\}$$

Accordingly, to test H_o, we construct

$$\underline{x}_t = \begin{bmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{bmatrix} = \begin{bmatrix} y_{1t} - 2y_{2t} + y_{3t} \\ y_{2t} - 2y_{3t} + y_{4t} \\ y_{3t} - 2y_{4t} + y_{5t} \end{bmatrix}$$

and if H_o be true, then $\mathcal{E}(\underline{x}_t) = \underline{0}$ for all t. The vector $\underline{x}_t$ are

$$\underline{x}_1' = (-0.4 \quad +0.3 \quad -0.1)$$
$$\underline{x}_2' = (+0.3 \quad +0.0 \quad -0.8)$$
$$\underline{x}_3' = (+0.6 \quad -0.4 \quad +0.1)$$
$$\underline{x}_4' = (-0.1 \quad -0.4 \quad +0.1)$$
$$\underline{x}_5' = (-0.4 \quad +0.1 \quad -0.2)$$
$$\underline{x}_6' = (-0.3 \quad +0.5 \quad -0.5)$$
$$\underline{x}_7' = (-0.6 \quad +0.0 \quad +0.4)$$
$$\underline{x}_8' = (-0.5 \quad -0.2 \quad +0.7)$$
$$\underline{x}_9' = (+0.1 \quad +0.3 \quad -0.5)$$
$$\underline{x}_{10}' = (+0.5 \quad -0.7 \quad -0.1)$$
$$\underline{x}_{11}' = (-0.5 \quad +0.2 \quad -0.1)$$
$$\underline{x}_{12}' = (-0.8 \quad +0.4 \quad -0.2)$$

thus $(\underline{\bar{x}} - \underline{\mu}_o)' = \frac{1}{12}(-2.1 \quad +0.1 \quad -1.2)$

and $C = \sum \underline{x}_t \underline{x}_t' - 12\underline{\bar{x}}\underline{\bar{x}}'$

$$= \begin{pmatrix} +2.63 & -1.15 & -0.40 \\ -1.15 & +1.49 & -0.70 \\ -0.40 & -0.70 & +1.92 \end{pmatrix} - \frac{1}{12}\begin{pmatrix} -2.1 \\ +0.1 \\ -1.2 \end{pmatrix}(-2.1 \quad +0.1 \quad -1.2)$$

$$= \frac{1}{12}\begin{pmatrix} +27.15 & -13.59 & -7.32 \\ -13.59 & +17.87 & -8.28 \\ -7.32 & -8.28 & +21.60 \end{pmatrix}$$

The test criterion (see 1.3.2) is then

$$\mathcal{L} = (-2.1 \quad +0.1 \quad -1.2)\begin{pmatrix} +27.15 & -13.59 & -7.32 \\ -13.59 & +17.87 & -8.28 \\ -7.32 & -8.28 & +21.60 \end{pmatrix}^{-1}\begin{pmatrix} -2.1 \\ +0.1 \\ -1.2 \end{pmatrix}$$

(note that the factors of 12 and 1/12 cancel so that it is unnecessary, and inadvisable computationally to divide by 12 where indicated to obtain $\mathcal{L}$).

The quantity $\mathcal{L}$ is obtained by a Doolittle process as indicated below

<u>Doolittle Proceedure</u>

← 12 C		→	$12(\bar{\bar{x}}-\mu_o)$	Check
+12.15	-13.59	-7.32	-2.1	+4.14
+27.15	-13.59	-7.32	-2.1	+4.14
1.	-0.50055248	-0.26961325	-0.07734806	+0.15248618
	+17.87	-8.28	+0.1	-3.90
	+11.06749180	-11.94400440	-0.95116013	-1.82771281
	1.	-1.07920062	-0.08594179	-0.16514245
		+21.60	-1.2	+4.8
		+6.73641124	-2.79268033	+3.94373014
		1.	-0.41456500	+0.58543488

$$\mathcal{L} = (-2.1)(-0.07734806) + (-0.95116013)(-0.08594179)$$
$$+ (-2.79268033)(-0.41456500) = 1.3978.$$

Critical regions for $\mathcal{L}$

size 0.001 : 4.63

size 0.005 : 2.91

size 0.01 : 2.33

size 0.025 : 1.69

size 0.05 : 1.29 .

The hypothesis of a linearly decreasing time to complete task over the five days would be rejected at a 5% level.

It is interesting to observe that had the test been performed on any consecutive <u>three</u> days only (which would lead to the following t^2-test criteria

$$\text{Days } 1,2,3 \quad \frac{12\left(\frac{-2.1}{12}\right)^2}{\frac{27.15}{12\times 11}} = 1.787 \qquad \text{refer to } F_{1:11}$$

$$\text{Days } 2,3,4 \quad \frac{12\left(\frac{0.1}{12}\right)^2}{\frac{17.87}{12\times 11}} = 0.006 \qquad \text{refer to } F_{1:11}$$

$$\text{Days } 3,4,5 \quad \frac{12\left(\frac{-1.2}{12}\right)^2}{\frac{21.60}{12\times 11}} = 0.733 \qquad \text{refer to } F_{1:11}\ )$$

none of the results would show as significant at the 5% level. The three $F_{1:11}$ ratios are of course correlated so that it would be difficult to give a significance level to any one in the presence of the other two; nevertheless, the value $F_{1:11}^{(.05)}$

is 4.84 which greatly exceeds any of the three values.

Considering the significance of $\mathcal{L}$ against the non-significance of the three $F_{1:11}$ ratios in the context of the data, we are not too surprised. A better straight line fit to three days in inevitably available as against a fit to five days. Considered out of context though, the data makes an interesting point:

> Suppose the 12 vector observations $\{\underline{x}_j\}_1^{12}$ are available from the normal trivariate density; it is required to test (H_o) that the mean vector of the normal density is a vector of zeros. The value of the test criterion is $\mathcal{L}$ = 1.3978 which is significant at between the 2.5% and 5% level so that H_o is rejected. Three separate tests performed on the component elements of $\{\underline{x}_j\}_1^{12}$, however, prove non-significant individually.

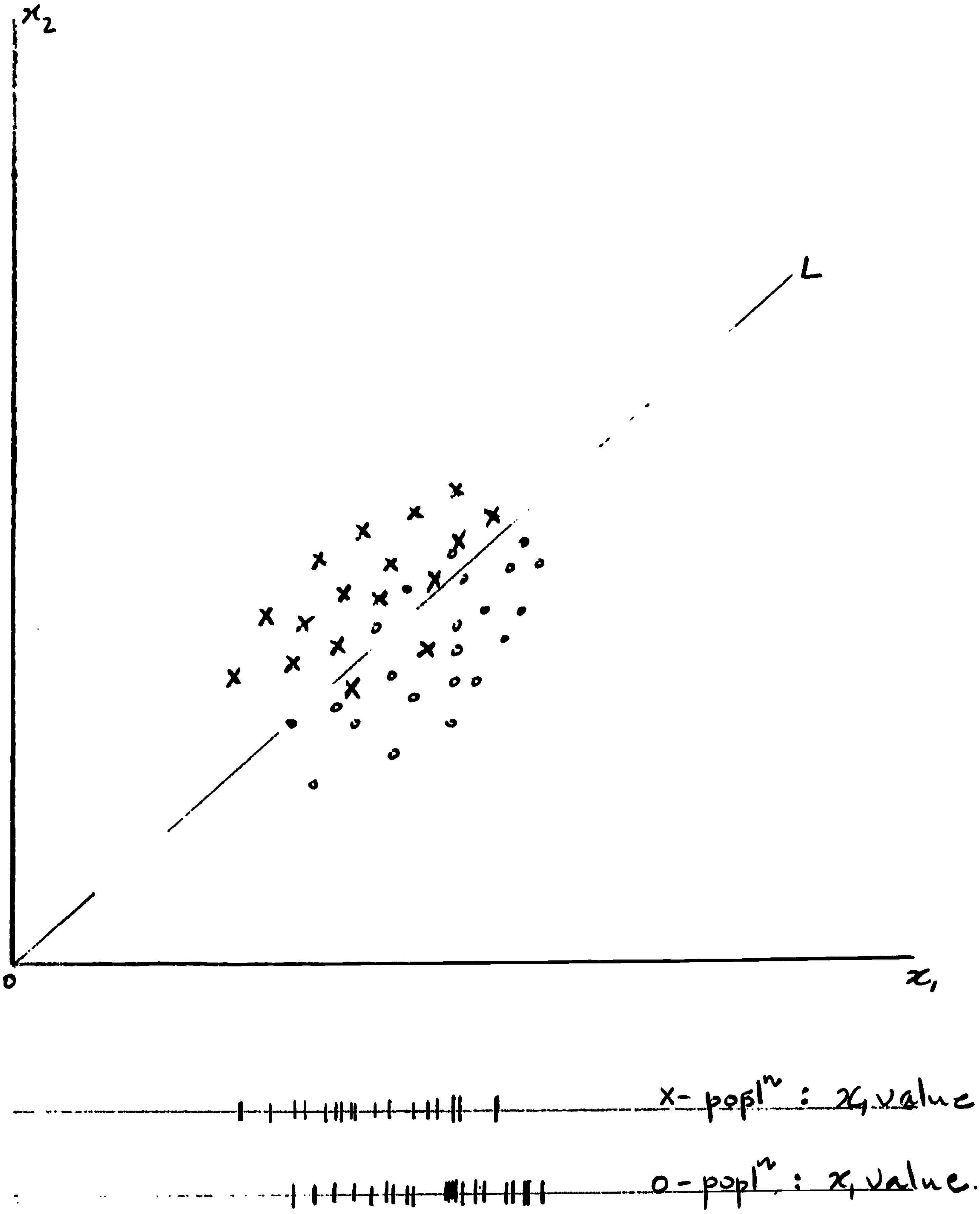
This kind of situation is illustrated diagrametrically in the next section.

11.6 A diagramatic illustration of the comparison of two population means in the bivariate case.

The following diagram is a dot diagram of two bivariate populations

x: represents a member of population 1

o: represents a member of population 2.


x2
L
0
x1
x- popl^n : x1 value
o- popl^n : x1 value.

A glance at the bivariate diagram is sufficient to indicate that the two populations are different in respect of their mean vector. All (or nearly all) the x-points lie inside an ellipse essentially above the line OL, nearly all the o-points lie inside an ellipse essentially below the line OL. Even without resort to multivariate techniques and an appropriate test, we would be prepared to pronounce the two populations different. If we tried to assess the populations in respect of one measurement alone (x_1, say), then we would be looking at the projection of all points (x's and o's) onto the abscissa of the diagram. It is noted that there is considerable intermingling of the projected points and one would be hard put to decide whether these points came from a population with the same $\mathcal{E}(x_1)$ or not. Similar remarks apply to the projection onto the ordinate of the diagram when the values of $\mathcal{E}(x_2)$ are being compared. It is clear then that the bivariate comparison is much more decisive than two univariate comparisons. Of course, the points have been chosen to illustrate the advantage of a bivariate test (in general, a multivariate test) as dramatically as possible; nevertheless, in cases where $\underline{\mu}_1$ and $\underline{\mu}_2$ vector means for two m-variate populations, are such that corresponding elements of $\underline{\mu}_1$ and $\underline{\mu}_2$ differ by only small quantities, then it may well happen that each of the m univariate tests would individually prove non-significant (using the standard t-test), whereas a multivariate test (Hotelling's) would give a significant result.

CHAPTER 12

ANALYSIS OF DESIGN

12.1 Introduction

We deal now with the analysis of data which are collected from a "designed experiment." In this category would be included the latin square; randomized block; cross-classifications; factorial designs; and so forth. The analysis is only slightly more complicated in the multivariate case than in the univariate case. This slight complication arises from the fact that the percentage points of the F-distribution provide the critical regions for the test in the univariate situation where as in the multivariate situation no such handy table of percentage points exist. The level of significance of any given value of the test criterion in the multivariate case is quite quickly established, however; the formula being given in a later section.

It is assumed that the reader has a familiarity with univariate proceedure and so only one situation will be developed; this should suffice to demonstrate that the multivariate proceedure is a very simple modification of the univariate proceedure. We choose to investigate the two-way cross classification of which the randomized block is a well known example.

12.2 The two-way cross-classification; algebraic development.

In our general design, let us assume r rows (blocks for example) and c columns (treatments perhaps). Each vector of correlated observations $\underline{x}$ is classifiable according to a row and a column. Supposing replicated vector observations (n per cell) then we shall write $\underline{x}_{ijk}$ for the k-th replicate in the i-th row, j-th column ($i = 1, 2, \ldots, r$; $j = 1, 2, \ldots, c$; $k = 1, 2, \ldots, n$).

The observations within a given cell will have the same mean vector which will, however, vary from cell to cell. Accordingly, our most general supposition would be:

$$\mathcal{E}(\underline{x}_{ijk}) = \underline{\gamma}_{ij} \qquad \begin{cases} i = 1, 2, \ldots, r \\ j = 1, 2, \ldots, c \end{cases} .$$

Our first test (possible only if n>1), sometimes called the "test of additivity", alternatively called "test for zero interactions" can be stated algebraically as

$$H_0: \underline{\gamma}_{ij} = \underline{\mu} + \underline{\alpha}_i + \underline{\beta}_j \qquad \begin{cases} i = 1, 2, \ldots, r \\ j = 1, 2, \ldots, c \end{cases}$$

where $\underline{\mu}$, $\{\underline{\alpha}_i\}_1^r$ and $\{\underline{\beta}_j\}_1^c$ are unknown. The hypothesis H_0 expresses the belief that the difference in expected "yield" ($\underline{x}$) for any two cells in the same row is a function only of the column number of those cells and not of the common row number; that is

$$\underline{\gamma}_{ij} - \underline{\gamma}_{ig} \text{ is independent of } i.$$

The multivariate analysis table is:

Source	Matrix	Degrees of Freedom
Between rows	$nc \sum_i (\overline{\underline{x}}_{i..} - \overline{\underline{x}}_{...})(\overline{\underline{x}}_{i..} - \overline{\underline{x}}_{...})'$	r-1
Between columns	$nr \sum_j (\overline{\underline{x}}_{.j.} - \overline{\underline{x}}_{...})(\overline{\underline{x}}_{.j.} - \overline{\underline{x}}_{...})'$	c-1
Interaction	$n\sum_{ij} (\overline{\underline{x}}_{ij.} - \overline{\underline{x}}_{i..} - \overline{\underline{x}}_{.j.} + \overline{\underline{x}}_{...})(\overline{\underline{x}}_{ij.} - \overline{\underline{x}}_{i..} - \overline{\underline{x}}_{.j.} + \overline{\underline{x}}_{...})'$	(r-1)(c-1)
Error	$\sum\sum\sum_{ijk} (\underline{x}_{ijk} - \overline{\underline{x}}_{ij.})(\underline{x}_{ijk} - \overline{\underline{x}}_{ij.})'$	rc(n-1)
Total	$\sum\sum\sum_{ijk} (\underline{x}_{ijk} - \overline{\underline{x}}_{...})(\underline{x}_{ijk} - \overline{\underline{x}}_{...})'$	nrc-1

The similarity of this table with the univariate table is strikingly obvious.

Designating the matrices as they occur in the table by M_R, M_C, M_I, M_E, and M_T respectively then

$$M_R + M_C + M_I + M_E = M_T \quad .$$

<u>To test for additivity</u>:

Construct $$\mathcal{L} = \frac{|M_E|}{|M_E+M_I|}$$

The use of the Doolittle proceedure is valuable in the computation of $|M_E|$ and $|M_E+M_I|$. Small values of $\mathcal{L}$ are significant of a contradiction to a hypothesis of additivity. In fact, if τ is an experimental value of $\mathcal{L}$ then it can be shown that, with

$$\rho_o = rc(n-1) - \tfrac{1}{2}(m+r+c-rc)$$

$$\pi_2 = \frac{m(r-1)(c-1)}{48\rho_o^2}(m^2+(r-1)^2(c-1)^2-5)$$

$$\pi_4 = \frac{m(r-1)(c-1)}{1920\rho_o^4}[3m^4+3(r-1)^4(c-1)^4+10m^2(r-1)^2(c-1)^2 -50m^2-50(r-1)^2(c-1)^2+159]$$

then

$$\Pr(-\rho_o \log \mathcal{L} \leq \tau) =$$

$$(1 - \pi_2 - \pi_4 + \tfrac{1}{2}\pi_2^2)\ \Pr\{\chi^2_{m(r-1)(c-1)} \leq \tau\}$$

$$+ (\pi_2 - \pi_2^2)\ \Pr\{\chi^2_{m(r-1)(c-1)+4} \leq \tau\}$$

$$+ (\pi_4 + \tfrac{1}{2}\pi_2^2)\ \Pr\{\chi^2_{m(r-1)(c-1)+8} \leq \tau\}$$

$$+ \text{ a term of order } \frac{1}{\rho_o^6} \quad .$$

It is noted that π_2 and π_4 are usually quite small and it is frequently only necessary to use

$$\Pr(-\rho_0 \log \mathcal{L} \leq \tau) = \Pr\{\chi^2_{m(r-1)(c-1)} \leq \tau\}.$$

12.3 An example in the test of additivity.

The following example relates to the performance of a group of retarded children. Any child can be categorized according to his I.Q. (broad group categories are used in our example) and according to the type of school attended. Three I.Q. broad categories are considered

Q_1: I.Q. of 60 or less

Q_2: I.Q. of 75 or less, but more than 60

Q_3: I.Q. in excess of 75 .

The types of school attended were three

S_1: public school (i.e. attended by the entire spectrum of I.Q.'s.)

S_2: special schools (i.e. attended by the somewhat slower child, however, individual attention is not the teaching method)

S_3: special schools (i.e. attended by the slower child, the emphasis on individual attention).

Five children were examined in each of the nine groups and judged on four tests; arithmetic, vocabulary, general science, constructive aptitude. The vectors listed are the scores on the tests, top-to-bottom in the order given above. [The scores themselves have been adjusted so that, in each test, the average score in the (Q_2,S_2)-group is approximately 75; this helps in getting a feel for the data, for example we observe that the second line in the (Q_3,S_1)-group is in the 80's--higher than "average" and the fourth line of the (Q_1,S_3)-group is rather higher also than the rest.]

TABLE OF OBSERVATIONS

		TYPE OF SCHOOL ATTENDED		
		S_1	S_2	S_3
I. Q. CLASSIFICATION	Q_1	$\begin{bmatrix}71\\74\\78\\67\end{bmatrix}\begin{bmatrix}70\\78\\79\\74\end{bmatrix}\begin{bmatrix}69\\74\\71\\72\end{bmatrix}\begin{bmatrix}69\\70\\73\\70\end{bmatrix}\begin{bmatrix}70\\77\\76\\70\end{bmatrix}$	$\begin{bmatrix}72\\72\\70\\81\end{bmatrix}\begin{bmatrix}76\\74\\70\\78\end{bmatrix}\begin{bmatrix}72\\71\\78\\77\end{bmatrix}\begin{bmatrix}74\\71\\76\\72\end{bmatrix}\begin{bmatrix}76\\72\\76\\81\end{bmatrix}$	$\begin{bmatrix}74\\70\\78\\86\end{bmatrix}\begin{bmatrix}81\\74\\78\\82\end{bmatrix}\begin{bmatrix}80\\79\\79\\89\end{bmatrix}\begin{bmatrix}80\\75\\72\\88\end{bmatrix}\begin{bmatrix}81\\72\\75\\88\end{bmatrix}$
	Q_2	$\begin{bmatrix}65\\76\\78\\72\end{bmatrix}\begin{bmatrix}65\\76\\77\\68\end{bmatrix}\begin{bmatrix}70\\79\\79\\66\end{bmatrix}\begin{bmatrix}69\\77\\76\\71\end{bmatrix}\begin{bmatrix}72\\79\\76\\71\end{bmatrix}$	$\begin{bmatrix}70\\74\\77\\70\end{bmatrix}\begin{bmatrix}72\\73\\72\\71\end{bmatrix}\begin{bmatrix}76\\73\\78\\73\end{bmatrix}\begin{bmatrix}71\\74\\76\\77\end{bmatrix}\begin{bmatrix}78\\77\\75\\73\end{bmatrix}$	$\begin{bmatrix}77\\73\\73\\82\end{bmatrix}\begin{bmatrix}79\\74\\74\\81\end{bmatrix}\begin{bmatrix}76\\70\\79\\85\end{bmatrix}\begin{bmatrix}78\\70\\77\\82\end{bmatrix}\begin{bmatrix}83\\74\\70\\84\end{bmatrix}$
	Q_3	$\begin{bmatrix}72\\83\\74\\67\end{bmatrix}\begin{bmatrix}74\\82\\79\\69\end{bmatrix}\begin{bmatrix}69\\82\\76\\72\end{bmatrix}\begin{bmatrix}74\\84\\75\\67\end{bmatrix}\begin{bmatrix}72\\78\\75\\73\end{bmatrix}$	$\begin{bmatrix}75\\73\\78\\76\end{bmatrix}\begin{bmatrix}72\\78\\73\\81\end{bmatrix}\begin{bmatrix}73\\78\\71\\76\end{bmatrix}\begin{bmatrix}78\\82\\74\\78\end{bmatrix}\begin{bmatrix}77\\81\\71\\74\end{bmatrix}$	$\begin{bmatrix}79\\79\\75\\76\end{bmatrix}\begin{bmatrix}77\\73\\76\\83\end{bmatrix}\begin{bmatrix}86\\79\\78\\82\end{bmatrix}\begin{bmatrix}78\\77\\73\\82\end{bmatrix}\begin{bmatrix}82\\77\\79\\75\end{bmatrix}$

*Within a given cell there are five replicates of (4x1)-column vectors.

Table of Sums

	S_1	S_2	S_3	Sub Total
Q_1	$\begin{bmatrix}349\\373\\377\\353\end{bmatrix}$	$\begin{bmatrix}370\\360\\370\\389\end{bmatrix}$	$\begin{bmatrix}396\\370\\382\\433\end{bmatrix}$	$\begin{bmatrix}1115\\1103\\1129\\1175\end{bmatrix}$
Q_2	$\begin{bmatrix}341\\387\\386\\348\end{bmatrix}$	$\begin{bmatrix}367\\371\\378\\364\end{bmatrix}$	$\begin{bmatrix}393\\361\\373\\414\end{bmatrix}$	$\begin{bmatrix}1101\\1119\\1137\\1126\end{bmatrix}$
Q_3	$\begin{bmatrix}361\\409\\379\\348\end{bmatrix}$	$\begin{bmatrix}375\\392\\367\\385\end{bmatrix}$	$\begin{bmatrix}402\\385\\381\\398\end{bmatrix}$	$\begin{bmatrix}1138\\1186\\1127\\1131\end{bmatrix}$
Sub Totals	$\begin{bmatrix}1051\\1169\\1142\\1049\end{bmatrix}$	$\begin{bmatrix}1112\\1123\\1115\\1138\end{bmatrix}$	$\begin{bmatrix}1191\\1116\\1136\\1245\end{bmatrix}$	$\begin{bmatrix}3354\\3408\\3393\\3432\end{bmatrix}$

Before computing tables of cross-products, we subtract 60 from every element (for computational convenience).

Table of Within Cell Cross Products

$$
\begin{array}{cccc}
 & S_1 & S_2 & S_3 \\
Q_1 & \begin{bmatrix} 483 & 620 & 764 & 515 \\ 620 & 1105 & 1150 & 788 \\ 764 & 1150 & 1231 & 814 \\ 515 & 788 & 814 & 589 \end{bmatrix} & \begin{bmatrix} 996 & 846 & 976 & 1248 \\ 846 & 726 & 826 & 1075 \\ 976 & 826 & 1036 & 1224 \\ 1248 & 1075 & 1224 & 1639 \end{bmatrix} & \begin{bmatrix} 1878 & 1366 & 1565 & 2554 \\ 1366 & 1026 & 1153 & 1875 \\ 1565 & 1153 & 1378 & 2171 \\ 2554 & 1875 & 2171 & 3569 \end{bmatrix} \\
Q_2 & \begin{bmatrix} 375 & 731 & 706 & 391 \\ 731 & 1523 & 1497 & 830 \\ 706 & 1497 & 1486 & 818 \\ 391 & 830 & 818 & 486 \end{bmatrix} & \begin{bmatrix} 945 & 964 & 1048 & 861 \\ 964 & 1019 & 1107 & 911 \\ 1048 & 1107 & 1238 & 1003 \\ 861 & 911 & 1003 & 848 \end{bmatrix} & \begin{bmatrix} 1759 & 1149 & 1327 & 2121 \\ 1149 & 761 & 865 & 1386 \\ 1327 & 865 & 1115 & 1669 \\ 2121 & 1386 & 1669 & 2610 \end{bmatrix} \\
Q_3 & \begin{bmatrix} 761 & 1334 & 968 & 572 \\ 1334 & 2397 & 1722 & 1025 \\ 968 & 1722 & 1263 & 761 \\ 572 & 1035 & 761 & 492 \end{bmatrix} & \begin{bmatrix} 1151 & 1398 & 1008 & 1262 \\ 1398 & 1742 & 1205 & 1564 \\ 1008 & 1205 & 931 & 1143 \\ 1262 & 1564 & 1143 & 1473 \end{bmatrix} & \begin{bmatrix} 2134 & 1756 & 1677 & 1993 \\ 1756 & 1469 & 1379 & 1650 \\ 1677 & 1379 & 1335 & 1575 \\ 1993 & 1650 & 1575 & 1978 \end{bmatrix}
\end{array}
$$

Matrix of total variation (M_T) [degrees of freedom: 44]

$$
\frac{1}{45}\begin{bmatrix} 43974 & -5652 & -1467 & 39537 \\ -5652 & 28296 & 36 & -18576 \\ -1467 & 36 & 15336 & -4266 \\ 39537 & -18576 & -4266 & 79956 \end{bmatrix}
$$

Matrix of Error variation (M_E) [degrees of freedom: 36]

$$
\frac{1}{5}\begin{bmatrix} 1324 & 107 & 3 & -167 \\ 107 & 1110 & -174 & -55 \\ 3 & -174 & 1412 & -194 \\ -167 & -55 & -194 & 1502 \end{bmatrix}
$$

Matrix of column (schooling) variation (M_C) [degrees of freedom: 2]

$$\frac{1}{45}\begin{bmatrix} 29562 & -10779 & -828 & 41322 \\ -10779 & 4974 & 1413 & -15231 \\ -828 & 1413 & 1206 & 1332 \\ 41322 & -15231 & 1332 & 57786 \end{bmatrix}$$

Matrix of Row (I.Q.) variation (M_R) [degrees of freedom: 2]

$$\frac{1}{45}\begin{bmatrix} 2094 & 4164 & -528 & -141 \\ 4164 & 11634 & -708 & -4101 \\ -528 & -708 & 168 & -354 \\ -141 & -4101 & -354 & 4362 \end{bmatrix}$$

Matrix of interaction variation (M_I) [degrees of freedom: 4]
[obtained by subtraction]

$$\frac{1}{45}\begin{bmatrix} 402 & 0 & -138 & -3147 \\ 0 & 1698 & 897 & 811 \\ -138 & 897 & 1254 & -3498 \\ -3147 & 811 & -3498 & 4290 \end{bmatrix}$$

Before proceding to the construction of the test statistic it is interesting to compare the matrices $\frac{1}{36} M_E$ and $\frac{1}{4} M_I$ in respect of their diagonal terms; this would correspond to four individual tests of additivity of effects on arithemtic score

additivity of effects on vocabulary score

additivity of effects on general science score

additivity of effects on constructive attitude score.

The F ratios are respectively: $\frac{402}{1324}$; $\frac{1698}{1110}$; $\frac{1254}{1412}$; $\frac{4290}{1502}$.

Only the last of these is significant at the 5% level when considered as an individual test.

The next step is to perform a Doolittle on M_E and on M_E+M_I to obtain the determinants of each.

Doolittle on $5M_E$

		$5M_E$		Check
1324	107	3	-167	1267
1324	107	3	-167	1267
1	0.080815	0.002265	-0.126132	0.956948
	1110	-174	-55	988
	1101.3528	-174.2424	-41.5039	885.6066
	1	-0.158207	-0.037684	0.804108
		1412	-194	1047
		1384.4268	-200.1878	1184.2389
		1	-0.144599	0.855400
			1502	
			1450.4250	

$$|M_E| = (1324)(1101.3528)(1384.4268)(1450.4250)/5^4$$

$$= 0.468489 \times 10^{10} \quad .$$

Doolittle on $45(M_E+M_I)$

	$45(M_E+M_I)$			Check
12318	963	-111	-4650	8520
12318	963	-111	-4650	8520
1	0.078178	-0.009011	-0.377496	0.691670
	11688	-669	316	12298
	11612.715	-660.322	679.529	11631.922
	1	-0.056861	0.058515	1.001653
		13962	-5244	7938
		13923.453	-5247.263	8676.189
		1	-0.376865	0.623134
			17808	
			14035.371	

$$|M_E+M_I| = (12318)(11612.715)(13923.453)(14035.371)/(45)^4$$

$$= 0.68170 \times 10^{10} \quad .$$

Finally

$$\mathcal{L} = \frac{|M_E|}{|M_E+M_H|} = 0.68724 \quad .$$

To establish the significance level of this experimental value we require $\Pr\{\mathcal{L} \leq 0.68672\}$. Now m=4; r=c=3; n=5 so that

$$\rho_o = 36 - \tfrac{1}{2}(4+3+3-9) = 35\tfrac{1}{2}$$

$$\pi_2 = \frac{4\cdot2\cdot2}{48\rho_o^2}.4^2+2^2\cdot2^2-5) = 0.007141$$

$$\pi_4 = \frac{4\cdot2\cdot2}{1920\rho_o^{1}}\cdot(3\cdot4^4+3\cdot2^4\cdot2^4+10\cdot4^2\cdot2^2\cdot2^2-50\cdot4^2-50\cdot2^2\cdot2^2+159)$$

$$= 0.87066 \times 10^{-6}$$

Clearly π_2^2 and π_4 can be neglected in the formula for the significant level. We have then

$$\Pr\{\mathcal{L} \leq 0.68724\} = \Pr\{-\log_e \mathcal{L} \geq 0.3758\}$$
$$= \Pr\{-35\tfrac{1}{2}\cdot\log_e \mathcal{L} \geq 13.34\}$$
$$= (0.992859)\Pr\{\chi^2_{16} \geq 13.34\} + (0.007141)\Pr\{\chi^2_{20} \geq 13.34\}$$
$$= 0.649 \quad .$$

A value of $\mathcal{L}$ less than or equal to the observed value of 0.68672 could be obtained approximately 65% of occassions by chance variation; there is no evidence to contradict the hypothesis that the effects are additive.

Since the so-called interaction term is not significantly different from zero, it would be meaningful to test the main effects. Suppose it is required to test that only the second (vocabulary) and fourth (constructive aptitude) scores differ with I.Q. level. A check on diagonal elements of row matrix versus error matrix suggests that the least contribution is due to the third score; the next to the first score, the next to the fourth score and the most contribution from the second score. We reorder our elements within the vector to conform with this contribution order. The listing of scores now becomes

$$\begin{bmatrix} \text{general science} \\ \text{arithmetic} \\ \text{constructive aptitude} \\ \text{vocabulary} \end{bmatrix} \quad .$$

Rearranging M_R and M_E to conform with this ordering we have

$$M_E = \frac{1}{5}\begin{bmatrix} 1412 & 3 & -194 & -174 \\ 3 & 1324 & -167 & 107 \\ -194 & -167 & 1502 & -55 \\ -174 & 107 & -55 & 1110 \end{bmatrix}$$

$$M_R = \frac{1}{45}\begin{bmatrix} 168 & -528 & -354 & -708 \\ -528 & 2094 & -141 & 4164 \\ -354 & -141 & 4362 & -4101 \\ -708 & 4164 & -4101 & 11634 \end{bmatrix} .$$

We perform forward Doolittles on each of M_E and $M_E + M_R$ =

$$\begin{bmatrix} 12876 & -501 & -2100 & -2274 \\ -501 & 14010 & -1644 & 5127 \\ -2100 & -1644 & 17880 & -4596 \\ -2274 & 5127 & -4596 & \text{[illegible]}1624 \end{bmatrix} .$$

Doolittle on $5M_E$

$5M_E$				Check
1412	3	-194	-174	1047
1412	3	-194	-174	1047
1	0.002124	-0.137393	-0.123224	0.741501
	1324	-167	107	1267
	1323.9936	-166.5878	107.3697	1264.7755
	1	0.125822	0.081095	0.955273
		1502	-55	1086
		1454.3853	-65.3970	1388.9880
		1	-0.044965	0.955034
			1110	
			1076.9104	

$$|M_E| = \frac{1}{5^4}(1412)(1323.9936)(1454.3853)(1076.9104)$$

$$= 0.468489 \times 10^{10} .$$

Doolittle on $45(M_E+M_C)$

		$45(M_E+M_C)$		Check
12876	-501	-2100	-2274	8001
12876	-501	-2100	-2274	8001
1	-0.038909	-0.163094	-0.176607	0.621385
	14010	-1644	5127	16992
	13990.5066	-1725.7101	5038.5199	17303.3154
	1	-0.123348	0.360138	1.236789
		17880	-4596	9540
		17324.6397	-4345.3809	12979.2541
		1	-0.250820	0.749179
			21624	
			16693.9248	

	d_{1j} diagonal terms of Doolittle on M_E	d_{2j} diagonal terms of Doolittle on M_E+M_R	F-ratio*	d.f.
j = 1	282.400	286.133	0.2379	2:36
2	264.799	310.900	3.0467	2:35
3	290.877	384.992	5.5004	2:34
4	215.382	370.976	11.9197	2:33

$$*\text{F-ratio} = \frac{(d_{2j}-d_{1j})/\text{degrees of freedom of } M_R}{d_{1j}/1-j + \text{degrees of freedom of } M_E} .$$

Of these ratios, the first and second (j=1,2) are not significant at the 5% level. The third, considered as a single test, would be significant at the 1% level (but not at the 0.5% level); the fourth figure is highly significant.

It seems reasonable to infer that, within the range considered, the performance in general science and arithmetic is relatively unaffected by the level of I.Q.; however, the individual will perform significantly differently in constructive aptitude and vocabulary.

The j-th F-ratio measures the contribution to significance over and above the contribution of the 1st, 2nd, ..., (j-1)-th variable. The F-ratio of 0.2379 measures the contribution to significance of the difference between the first element in the vector

general science	score
arithmetic	score
constructive aptitude	score
vocabulary	score

between the I.Q. levels. The figure is so small that we elect to state that no differences exist; an examination of the table of sums (third element in sub-totals for rows: 1129; 1137; 1127) fortifies our belief in the non-existence of any difference. The F-ratio 3.0467 is not significant at the 5% level for single tests; this ratio corresponds to a comparison of the arithmetic scores for which the subtotals are 1115; 1101; 1138. The F-ratio 5.5004 (constructive aptitude scores) corresponds to the subtotal figures 1175; 1126; 1131. This last F-ratio is rather border-line and one may hesitate to assert that the difference is significant (bearing in mind that several tests have been made and that the data has been ordered the "F-ratio" does not have exactly the F-distribution). The final F-ratio of 11.9197 corresponds to the vocabulary scores for which the row (I.Q.) subtotals are 1103; 1119; 1186. We assert with some confidence that the vocabulary score differs significantly between the I.Q. groups within the range considered.

CHAPTER 13

ANALYSIS OF REGRESSION

13.1 Introduction

In the analysis of regression we lack the simplicity of computation usually found in the analysis of design; to compensate for this however is the fact that the so-called "design matrix" is of full rank, the result of which is that the square matrices of our normal equations do indeed have inverses. We look in this chapter at two problems (which actually are essentially the same)

(a) the curvilinear regression on a single concommitant variable

(b) the multiple regression on several concommitant variables.

13.2 Curvilinear regression (general case).

Observed are n_i $(m\times 1)$ vector variates at time (or temperature or any other concommitant variable) t_i. We denote the observations by $\{\underline{y}_{ij}\}_{i=1,j=1}^{k\ \ n_i}$ so that k different temperatures are involved.

Let $u_j(t)$ be a polynomial in t of degree j $(j=0,1,2,\ldots)$; these polynomials are arbitrary except in special cases we will usually adopt the system $u_j(t) = t^j$. Our objective is to test the hypothesis that each element in the vector $\mathcal{E}(\underline{y}_{ij})$ is a polynomial in t of degree no more than s.

We may write our model:

13.2.1 $$\mathcal{E}(\underline{y}_{ij}) = \underline{\beta}_o u_o(t_i)+\underline{\beta}_1 u_1(t_i)+\ldots+\underline{\beta}_s u_s(t_i)+\underline{d}_s(t_i)$$

where the $\sum_{j=0}^{s} \underline{\beta}_j \underline{u}_j(t_i)$ represents a general polynomial of degree s and where $\underline{d}_s(t_i)$ represents a "departure from s-th degree polynomial." The analysis table again has the striking resemblance to the univariate table, squares in the latter become the product of the column vector by its transpose for the latter. The table is:

Source	Matrix	Degrees of Freedom
Due to polynomial of degree s	$\sum_{i=1}^{k} \hat{\underline{Y}}_i \hat{\underline{Y}}_i' n_i$	s+1
About Polynomial of degree s	$\sum_{i=1}^{k} (\bar{\underline{Y}}_{i.} - \hat{\underline{Y}}_i)(\bar{\underline{Y}}_{i.} - \hat{\underline{Y}}_i)' n_i$	k-s-1
Error	$\sum_{i=1}^{k} \sum_{j=1}^{n_i} (\underline{y}_{ij} - \bar{\underline{Y}}_{i.})(y_{ij} - \bar{\underline{Y}}_{i.})'$	N-k
Total	$\sum_{i=1}^{k} \sum_{j=1}^{n_i} (\underline{y}_{ij} \underline{y}_{ij}')$	N

In the table $\hat{\underline{y}}_i$ is the best estimate of $\mathcal{E}(\underline{y}_{ij})$ on the assumption that H_o is true and is obtained via

$$\hat{\underline{Y}}_i = \hat{\underline{\beta}}_o u_o(t_i) + \hat{\underline{\beta}}_1 u_1(t_i) + \dots + \hat{\underline{\beta}}_s u_s(t_i)$$

where

$$\{\hat{\underline{\beta}}_\gamma\}_{\gamma=1}^{s} \quad \text{are the solutions of}$$

$$\sum_{i=1}^{k} n_i \bar{\underline{Y}}_{i.} u_\gamma(t_i) = \sum_{\ell} \hat{\underline{\beta}}_\ell [\sum_i u_\ell(t_i) u_\gamma(t_i) n_i]$$

and $$\bar{\underline{Y}}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} \underline{y}_{ij}$$

$$\bar{\underline{Y}}_{..} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\underline{Y}_{ij} \qquad (N = \sum_i n_i) \quad .$$

Clearly the stickiest part of the analysis is the solution for the $\{\underline{\beta}_\gamma\}_{\gamma=0}^{s}$, however this is accomplished routinely through a Doolittle computation.

Let G be the matrix ($m\times\overline{s+1}$) whose γ-th column is

$$\sum_i n_i\bar{\underline{Y}}_{i.}u_\gamma(t_i) \qquad (\gamma = 0, 1, \ldots, s)$$

and A be the matrix (symmetric $\overline{s+1}\times\overline{s+1}$) whose (ℓ,γ)-th element is

$$\sum_i n_i u_\ell(t_i)u_\gamma(t_i) \qquad \begin{matrix}\ell\\ \gamma\end{matrix} = 0, 1, 2, \ldots, s \quad .$$

If $\hat{B}$ is the ($m\times\overline{s+1}$) matrix $\hat{B} = (\hat{\underline{\beta}}_o, \hat{\underline{\beta}}_1, \ldots, \hat{\underline{\beta}}_s)$ then our system of equations on the $\{\hat{\underline{\beta}}_\gamma\}$ can be written as

$$G = \hat{B}A$$

with G and A known. The problem of finding $\hat{B}$ is dealt with in Chapter 10. Having got $\hat{B}$, the $\hat{Y}_i$ are easy to calculate and thence the entries in the analysis table.

Labelling the four matrices in the table, top to bottom, by C_D, C_A, C_E, and C_T respectively then the test function is

$$\mathcal{L} = \frac{|C_A|}{|C_E+C_A|}$$

small values of $\mathcal{L}$ are significant.

To assess the significance of an experimental value of we set

$$\nu_1 = k-s-1$$

$$\nu_2 = N-k$$

$$\rho_o = \nu_2 - \tfrac{1}{2}(m+1-\nu_1)$$

$$\pi_2 = \frac{m\nu_1}{48\rho_o^2}(m^2+\nu_1^2-5)$$

$$\pi_4 = \frac{m\nu_1}{1920\rho_o^4}(3m^4+3\nu_1^4+10m^2\nu_1^2-50m^2-50\nu_1^2+159)$$

and note that

$$\Pr(-\rho_o \log \mathcal{L} \le \tau) = (1-\pi_2-\pi_4-\pi_2^2)\Pr(\chi_{m\nu_1} \le \tau) + (\pi_2-\pi_2^2)\Pr(\chi^2_{m\nu_1+4} \le \tau)$$
$$+ (\pi_4+\pi_2^2)\Pr(\chi^2_{m\nu_1+8} \le \tau) + \text{a term of order } \rho_o^{-6} .$$

13.3 Curvilinear regression; no replication.

In the case of no replication ($n_i=1$), the "error matrix" of the previous section is identically zero; we need some kind of estimate of the common dispersion matrix of the $\underline{y}_i$. To consolidate ideas, let us set up our model again for this case of no replication; the suffix "j" is now dropped since j can only take the value 1. The model is

$$\mathcal{E}(\underline{y}_i) = \underline{\beta}_o u_o(t_i) + \underline{\beta}_1 u_1(t_i) + \dots + \underline{\beta}_s u_s(t_i) + d_s(t_i).$$

If it should turn out that s is two small to represent the degree of the polynomial which expresses the behaviour of $\mathcal{E}(\underline{y}_i)$ then the "about polynomial of degree s" of the previous section would have proved significant; on the other hand, had the true degree of the polynomial been s or less then the "about polynomial of degree s" would itself have been a measure of the dispersion matrix of a vector observation. We use this last statement to deal with this unreplicated data. It is required to test the hypothesis (H_o) that the elements of $\mathcal{E}(\underline{y}_i)$ lie on a polynomial of order not exceeding s. In preparing an analysis table it is to be remembered that H_o may be false in which case the degree of the polynomial will indeed be greater than s; in this event the "about polynomial of degree s" will not be a measure of the common dispersion matrix but will be confounded with variations of the true polynomial about the fitted s-th order polynomial.

We find ourselves forced to make some pronouncement concerning the true order of the polynomial and state (for some $\ell > 0$) that we believe that the order of the regression polynomial is not in excess of $s+\ell$. For example, we may wish to test that the regression is quadratic (s=2) feeling that (H_1) if a quadratic is inadequate then a cubic must surely be a polynomial of sufficiently high order to represent the data; we might feel safe in taking $\ell=1$ or to be "on the safe side" it might be preferred that ℓ be set equal to 3 As in the previous section, let G be the $(m \times \overline{s+1})$ matrix whose γ-th column is:

$$\sum \underline{y}_i u_\gamma(t_i) \qquad \gamma = 0, 1, \ldots, s$$

and A be the matrix (symmetric, $\overline{s+1} \times \overline{s+1}$) whose (δ,γ)-th element is

$$\sum_i u_\delta(t_i) u_\gamma(t_i) \qquad \begin{Bmatrix}\delta \\ \gamma\end{Bmatrix} = 0, 1, 2, \ldots, s \quad ,$$

then if $\hat{B}$ is the $(m \times \overline{s+1})$ matrix $\hat{B} = (\hat{\underline{\beta}}_o, \hat{\underline{\beta}}_1, \ldots, \hat{\underline{\beta}}_s)$, then $\hat{B}$ is the solution of (see Chapter 10)

$$G = \hat{B}A \quad .$$

If now

$$\hat{\underline{y}}_i = \hat{\underline{\beta}}_o u_o(t_i) + \hat{\underline{\beta}}_1 u_1(t_i) + \ldots + \hat{\underline{\beta}}_s u_s(t_i)$$

then $\sum_{i=1}^{k} \hat{\underline{y}}_i \hat{\underline{y}}_i'$ is the "due to polynomial of degree s" and

$\sum_{i=1}^{k} (\underline{y}_i - \hat{\underline{y}}_i)(\underline{y}_i - \hat{\underline{y}}_i)'$ is a measure of the variation not accounted for by the s-th order polynomial; this will be a measure of if and only if H_o is true.

Let now G^* be the $(m \times \overline{s+\ell+1})$ matrix whose γ-th column is

$$\sum \underline{y}_i u_\gamma(t_i) \qquad \gamma = 0, 1, 2, \ldots, s+\ell$$

so that G is the first s+1 columns of G^* and let A be the matrix (symmetric, $\overline{s+\ell+1} \times \overline{s+\ell+1}$) whose (δ,γ)-th element is

$$\sum_i u_\delta(t_i) u_\gamma(t_i) \qquad \left\{\begin{matrix}\delta \\ \gamma\end{matrix}\right. = 0, 1, 2, \ldots, s+\ell$$

then if $\hat{B}^*$ is the $(m \times \overline{s+\ell+1})$ matrix $\hat{B}^* = (\hat{\underline{\beta}}_o^*, \hat{\underline{\beta}}_1^*, \ldots, \hat{\underline{\beta}}_{s+\ell}^*)$ then $\hat{B}^*$ is the solution (see Chapter 10)

$$G = \hat{B}^* A \quad .$$

Now set

$$\hat{\underline{y}}_i^* = \hat{\underline{\beta}}_o^* u_o(t_i) + \hat{\underline{\beta}}_1^* u_1(t_i) + \ldots + \hat{\underline{\beta}}_{s+\ell}^* u_{s+\ell}(t_i)$$

then $\sum_{i=1}^{k} (\underline{y}_i - \hat{\underline{y}}_i^*)(\underline{y}_i - \hat{\underline{y}}_i^*)$ is a measure of the variation of the $\underline{y}_i$ about a polynomial of degree $s+\ell$. If our choice of ℓ is not too small, this should be purely random fluctuation.

Our analysis table is:

Source	Matrix	Degrees of Freedom
Due to polynomial of degree s.	$\sum \hat{y}_i \hat{y}_i'$	s+1
Gain in fitting polynomial of degree s+ℓ	$\sum \hat{y}_i^* \hat{y}_i^{*\prime} - \sum \hat{y}_i \hat{y}_i'$	ℓ
About polynomial of degree s+ℓ	$\sum (y_i - \hat{y}_i^*)(y_i - \hat{y}_i^*)'$	k-ℓ-s-1
Total	$\sum_i y_i y_i'$	k

If the matrices listed above are, top to buttom C_D, C_G, C_A and C_T then $C_D + C_G + C_A = C_T$. Our test function is

$$\mathcal{L} = \frac{|C_G|}{|C_G + C_A|}$$

and to assess the significance of $\mathcal{L}$, we set

$$\nu_1 = \ell$$

$$\nu_2 = k-\ell-s-1$$

$$\rho_o = \nu_2 - \tfrac{1}{2}(m+1-\nu_1)$$

$$\pi_2 = \frac{m\nu_1}{48\rho_o^2}(m^2+\nu_1^2-5)$$

$$\pi_4 = \frac{m\nu_1}{1920\rho_o^4}(3m^4+10m^2\nu_1^2+3\nu_1^4-50(m^2+\nu_1^2)+159)$$

then

$$\Pr(-\rho_o \log \mathcal{L} < \tau) = (1-\pi_2-\pi_4-\pi_2^2)\Pr(\chi^2_{m\nu_1}<\tau)+(\pi_2-\pi_2^2)\Pr(\chi^2_{m\nu_1+4}<\tau)$$

$$+(\pi_4+\pi_2^2)\Pr(\chi^2_{m\nu_1+8}<\tau)+\text{a term of order } \rho_o^{-6}.$$

CHAPTER 14

TEST OF HYPOTHESIS ON THE DISPERSION MATRIX

14.1 Introduction

We have seen in the preceding chapters how tests of hypothesis in the multivariate case are directly related to the corresponding test in the univariate case; the only essential difference is the distribution of the test criterion.

In the univariate case we have only a single (unknown) parameter, σ, expressing the standard deviation of the random error; in the multivariate case we have the array of parameters in V, the dispersion matrix of the vector observation $\underline{x}$. As a consequence we have the possibility of hypothesis concerning the elements of V which do not have a counterpart in the univariate methods and theory. In this chapter we shall be concerned with a test of independence between specified groups of elements of the observed variate $\underline{x}$.

14.2 The intra-independence of elements of $\underline{x}$.

Let us assume we have available the sequence of observations $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ randomly independently drawn from a normal population (or set of populations) with dispersion matrix (or dispersion matrices all equal to) V. For any given model for the mean vectors, $\mathcal{E}(\underline{x}_j)$ we can construct an error matrix, M_E say. Examples of such a construction are given in earlier chapters; in Chapter 12, M_E would be the error matrix in the cross-classification. If $\mathcal{E}(\underline{x}_j) = \underline{\mu}$, $j = 1,\ldots,n$, so that $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ are drawn from the same population (that is, same mean, same dispersion matrix) then

$$M_E = \sum_{j=1}^{n} (\underline{x}_j - \bar{\underline{x}}_.)(\underline{x}_j - \bar{\underline{x}}_.)' \qquad (\nu = n-1 \text{ below})$$

where $\bar{\underline{x}} = \frac{1}{n} \sum_j \underline{x}_j$.

Suppose the degrees of freedom associated with the error matrix is ν ($\nu = rc(n-1)$ in the example in Chapter 12), then we may require to test

$$H_o: \; v_{ij} = 0 \quad , \quad i \neq j \quad , \quad \begin{smallmatrix} i \\ j \end{smallmatrix} = 1, 2, \ldots, m$$

that is all off diagonal elements of V are zero. If $\underline{x}$ has the multivariate normal density the V diagonal is a necessary and sufficient condition for the mutual independence of <u>all</u> elements of $\underline{x}$.

The appropriate likelihood ratio test function is monotonic in

$$\mathcal{L} = |M_E| \Big/ \prod_{j=1}^{m} (M_E)_{jj}$$

where $(M_E)_{jj}$ is the j-th diagonal element of M_E. The null distribution of this criterion is unobtainable except for m=2 when $\mathcal{L}$ reduces to essentially the well known test for independence between two variates x_1 and x_2 (that is $\underline{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$).

For values of m exceeding two, we resort to an approximation to obtain the significance level of an observed value of $\mathcal{L}$. With

$$f = \tfrac{1}{2} m(m-1)$$

$$\rho_o = \nu - \frac{2m+5}{6}$$

then $\Pr[-\rho_o \log \mathcal{L} > \tau] = \Pr[\chi_f^2 > \tau] + O(\frac{1}{\rho_o^2})$.

By way of illustration we take the error matrix of the example in Chapter 12. [It is noted there that the off-diagonal elements in M_E are small compared with the diagonal terms, suggesting the possibility of mutual independence.

Now $$|M_E| = 0.468489 \times 10^{10}$$

and $$\prod_{j=1}^{4} (M_E)_{jj} = 0.498696 \times 10^{10}$$

so that $\mathcal{L} = 0.9394$; $f=6$; $\rho_o = 203/6$.

Now small values of $\mathcal{L}$ are significant and

$$\Pr[\mathcal{L} < 0.9394] = \Pr[\log \mathcal{L} < -0.0625]$$
$$= \Pr[-\rho_o \log \mathcal{L} > 2.11] = \Pr[\chi_6^2 > 2.11]$$
$$= 0.92 \quad .$$

There is every evidence therefore that the hypothesis of mutual independence of the scores discussed in Chapter 12 can be held to be true.

14.3 The inter-independence of two sets of variates.

Again we assume available the sequence of observations $\underline{x}_1$, $\underline{x}_2, \ldots, \underline{x}_n$ randomly and independently drawn from a normal population (or populations) such that the error matrix (M_E) has ν degrees of freedom. For convenience we shall drop the suffix E to M_E in this section writing simply M for the error matrix. If $\underline{x}$ is partitioned into two sets of elements $\underline{x}_{(1)}$ and $\underline{x}_{(2)}$ so that

$$\underline{x} = \begin{pmatrix} \underline{x}_{(1)} \\ \cdots \\ \underline{x}_{(2)} \end{pmatrix}$$

with p elements in the vector $\underline{x}_{(1)}$ and $q = m-p$ elements in the vector $\underline{x}_{(2)}$, then it may be required to test that every element of $\underline{x}_{(1)}$ is independent of every element of $\underline{x}_{(2)}$ (whilst still allowing that there may be intra-dependence in either or both partition vectors). If the dispersion matrix is V and we partiiton V as:

$$V = \begin{pmatrix} V_{11} & \vdots & V_{12} \\ \cdots & \vdots & \cdots \\ V_{21} & \vdots & V_{22} \end{pmatrix}$$

with V_{11} a pxp matrix and V_{22} a qxq matrix then the hypothesis of inter-independence can be statied as:

$$H_o: \ V_{12} = (0); \quad (V_{12} \text{ a matrix of all zeros}).$$

If M be partitioned in the same way as was V, that is

$$M = \begin{pmatrix} M_{11} & \vdots & M_{12} \\ \cdots & \vdots & \cdots \\ M_{21} & \vdots & M_{22} \end{pmatrix}; \qquad \begin{matrix} M_{11} \text{ is } p \times p; \ p+q = m \\ M_{22} \text{ is } q \times q \end{matrix}$$

then our test criterion is

$$\mathcal{L} = |M| / |M_{11}||M_{22}| \quad .$$

The distribution of $\mathcal{L}$ independent upon p and q and we have a readily available method of getting a critical region for $\mathcal{L}$ if p or q is either 1 or 2; otherwise we resort to a very adequate approximation.

Case p=1. (q=m-1).

$\dfrac{1-\mathcal{L}}{\mathcal{L}} \times \dfrac{\nu-q}{q}$ is distributed as $F_{q:\nu-q}$, (large values significant).

Case q=1. (p=m-1).

$$\frac{1-\mathcal{L}}{\mathcal{L}} \times \frac{\nu-p}{p} \text{ is distributed as } F_{p:\nu-p}. \text{ (large values significant).}$$

Case p=2. (q=m-2).

$$\frac{1-\mathcal{L}^{1/2}}{\mathcal{L}^{1/2}} \times \frac{\nu-q-1}{q} \text{ is distributed as } F_{2q:2(\nu-q-1)}. \text{ (large values significant).}$$

Case q=2. (p=m-2).

$$\frac{1-\mathcal{L}^{1/2}}{\mathcal{L}^{1/2}} \times \frac{\nu-p-1}{p} \text{ is distributed as } F_{2p:2(\nu-p-1)}. \text{ (large values significant).}$$

If both p and q exceed 2 then we set

$$\rho_o = \nu - \frac{m+1}{2}$$

$$\pi_2 = \frac{pq}{48\rho_o^2}(p^2+q^2-5)$$

$$\pi_4 = \frac{pq}{1920\rho_o^4}(3(p^4+q^4)+10p^2q^2-50(p^2+q^2)+159)$$

then

$$\Pr[-\rho_o \log \mathcal{L} > \tau] = (1-\pi_2-\pi_4+\tfrac{1}{2}\pi_2^2)\Pr[\chi^2_{pq} > \tau]$$

$$+ (\pi_2-\pi_2^2)\Pr[\chi^2_{pq+4}>\tau] + (\pi_4+\tfrac{1}{2}\pi_2^2)\Pr[\chi^2_{pq+8}>\tau]$$

$$+ \text{ terms of order } \rho_o^{-6}.$$

14.4 Equality of a number of dispersion matrices.

Suppose from each of k populations we have available an estimate of the dispersion matrix based on an "error-matrix" C_t with, say, ν_t degrees of freedom $(t=1,2,\ldots,k)$. The estimate of V_t, the dispersion matrix for the t-th population would be therefore $\frac{1}{\nu_t}C_t$. It is required to test

$$H_o: \; V_1 = V_2 = \ldots = V_k \quad .$$

The likelihood ratio test currently used is that derivable from the joint density of the $\{C_t\}_{t=1}^{k}$ rather than from the original (normally distributed) variates. This likelihood ratio is u where

$$\log u = (\tfrac{1}{2}m\sum_t \nu_t)\log(\sum\nu_t) - \tfrac{m}{2}\sum_t \nu_t \log\nu_t$$
$$+ \tfrac{1}{2}\sum_t \nu_t \log|C_t| - \tfrac{1}{2}(\sum\nu_t)\log|\sum_t C_t| \quad .$$

The distribution of u, or actually log u, can be approximated to by a χ^2-density. If we define

$$\rho_o = 1 - \frac{2m^2+3m-1}{6(m+1)(k-1)}\left\{\sum \frac{1}{\nu_t} - (\sum\nu_t)^{-1}\right\}$$

and define γ by:

$$48\rho_o^2 = m(m^2-1)(m+2)\left\{\sum \frac{1}{\nu_t^2} - (\sum\nu_t)^{-2}\right\}$$
$$-6m(m+1)(k-1)(1-\rho_o)^2$$

then with

$$f = \tfrac{1}{2}m(m+1)(k-1),$$

it can be shown that

$$\Pr\{-2\log u > \tau\} = (1-\gamma)\Pr\{\chi_f^2 > \tau\} + \gamma\Pr\{\chi_{f+4}^2 > \tau\}$$
$$+ \text{ terms of order } (\sum\nu_t)^{-3} \; .$$

Small values of u (and therefore large values of $-2\rho\log u$) are significant of departures from H_o.

CHAPTER 15

LINEAR DISCRIMINANTS

15.1 Introduction

The use of linear discriminants is quite wide spread in the problem of "reducing the dimension of the data." Suppose it is intended to make a detailed survey of, say, the general build and physical development of 15 year old male children according to their environment; urban, suburban, and rural. There are a large number of measurements which can be made on the human body: height; weight; waist measure, chest measure (exhaled and inhaled); length of leg, arm; distance around neck, head, calf, thigh; shoulder width--and so on. Each of these measurements is an indication of build and physical development which may vary according to the environment. However, for obvious reasons, it is undesirable to amass a superfluity of measurements on a large group of individuals.

There are basically two reasons why a particular measurement could be excluded from consideration:

(i) it is so highly related to another observation that it contributes little to our knowledge in the presence of this other variable.

As an example, shoulder width and chest girth are highly correlated variables. A casual glance at the human race is sufficient to indicate that the arms hang down from the shoulders touching the side of the rib cage in almost all cases so that we would intuitively guess that this high correlation exists. However, other variables (height and waist) have low correlation; it is noticeable that people of the same height differ quite considerable in their general build and in particular in the waist measurement.

(ii) a variable should also be considered for exclusion if it does not differ with, in our case, different environment, or if it differs so little as to be of little use in our investigation.

As an example, many measurements on the skeleton (particularly the skull) are no more variable across categories (suburban, urban, rural) than they are within each category and therefore, are unlikely to contribute much to our understanding of variations across these categories.

Rolf E. Bargmann (Part III of this contract) has given a detailed account of the use and application of discriminant functions so that in this volume we give only a resume.

15.2 Selection of significant variables.

Suppose we have k groups with, say, n_i vector observations in the i-th group:

Group I	Group II	...	Group i	...	Group k
$\underline{x}_{11},\underline{x}_{12},\ldots,\underline{x}_{1n_1}$	$\underline{x}_{21},\underline{x}_{22},\ldots,\underline{x}_{2n_2}$		$\underline{x}_{i1},\underline{x}_{i2},\ldots,\underline{x}_{in_i}$		$\underline{x}_{k1},\underline{x}_{k2},\ldots,\underline{x}_{kn}$

and we shall further suppose that each vector observation contains m elements, (possibly the anthropometric measurements discussed in section 15.1). The expectation of most elements of $\underline{x}_{ij}$ will change with group; some however (skull measurements perhaps) will not.

Rather than work with a vector of observations we choose to work with a scalar observation defined as a linear combination of the elements of the vector. Thus if $\underline{x}_{ij}$ is the j-th replicate in group i, we construct

$$z_{ij} = \underline{\alpha}'\underline{x}_{ij}, \qquad j = 1,2,\ldots,n_i;\ i=1,2,\ldots,k$$

for some, at the moment general $\underline{\alpha}$. The observation in group i are now $z_{i1}, z_{i2}, \ldots, z_{in_i}$ a set of n_i independent scalar observations.

Using the standard <u>univariate</u> techniques, we have

Source	Sum of Squares	Degrees of Freedom
Between group	$\sum_i (\bar{z}_{i.} - \bar{z}_{..})^2 n_i$	$k-1$
Within group	$\sum_i \sum_j (z_{ij} - \bar{z}_{i.})^2$	$\sum_{i=1}^{k} (n_i - 1)$

where

$$\bar{z}_{i.} = \frac{1}{n_i} \sum_j z_{ij} \qquad i=1,2,\ldots,k$$

$$\bar{z}_{..} = \frac{\sum_i \sum_j z_{ij}}{\sum_i n_i} \, .$$

Our usual test statistic expressing a difference between the groups is

$$\mathcal{F} = \frac{\frac{1}{k-1} \sum_i (\bar{z}_{i.} - \bar{z}_{..})^2 n_i}{\frac{1}{\sum_{i=1}^{k} (n_i - 1)} \sum_i \sum_j (z_{ij} - \bar{z}_{i.})^2}$$

referred to F-tables, degrees of freedom $\{k-1; \sum_i (n_i - 1)\}$.

The choice of $\underline{\alpha}$ is arbitrary at the moment but it seems reasonable to select that value of $\underline{\alpha}$ which maximizes $\mathcal{F}$. It is easily shown that $\underline{\alpha} = \underline{\alpha}_*$ must satisfy

$$(H - \theta E)\underline{\alpha}_* = \underline{0}$$

where

$$E = \sum_i \sum_j (\underline{x}_{ij}-\underline{\bar{x}}_{i.})(\underline{x}_{ij}-\underline{\bar{x}}_{i.})'$$

(the familiar "error matrix")

and

$$H = \sum_i (\underline{\bar{x}}_{i.}-\underline{\bar{x}}_{..})(\underline{\bar{x}}_{i.}-\underline{\bar{x}}_{..})'$$

(the familiar "hypothesis matrix").

θ is the maximal solution of $|H-\theta E| = 0$ and $\underline{\alpha}$ the associated eigen vector.

Of course with $\underline{\alpha}$ so selected the "between group" and "within group" sums of squares no longer have χ^2-distributions. However, we may say that $z_{ij} = \underline{\alpha}'\underline{x}_{ij}$ is the linear function of the elements of $\underline{x}_{ij}$ which discriminates best between the groups. If

$$\underline{x}_{ij} = \begin{bmatrix} {}_1x_{ij} \\ {}_2x_{ij} \\ \vdots \\ {}_mx_{ij} \end{bmatrix}$$

then we construct a table of correlations between

$$z_{ij} \text{ and } {}_qx_{ij} \quad , \qquad q = 1, 2, \ldots, m ,$$

that is, we compute

$$r_q = \frac{\sum_j (z_{ij}-\bar{z}_{i.})({}_qx_{ij}-{}_q\bar{x}_{i.})}{(\sum_j (z_{ij}-\bar{z}_{i.})^2)^{\frac{1}{2}}(\sum_j ({}_qx_{ij}-{}_q\bar{x}_{i.})^2)^{\frac{1}{2}}}$$

the inference is that a high value of r_q indicates that the elements ${}_qx_{ij}$ are "almost as good" as z_{ij} in discriminating between the groups.

The set of small correlations are noted ($r_q \leq 0.25$ sat) and it may be decided not to make the corresponding measure ${}_qx_{ij}$ in a large scale experiment.

This approach, it is admitted, is somewhat lacking in distributional justification; however, it does give us some idea of the relative rôles of the elements $\{{}_qx_{ij}\}_{q=1}^{n}$ in respect of their differences across the groups. Having decided that certain elements are of the non-contributing type, it is as well to use the step-down approach (section 12.3) as a check.