A NOTE ON THE DISTRIBUTION OF $D^2_{p+q}-D_p^2$ AND SOME COMPUTATIONAL ASPECTS OF D^2 STATISTIC AND DISCRIMINANT FUNCTION

By C. RADHAKRISHNA RAO
Statistical Laboratory, Calcutta

Introduction

In an earlier paper (Rao, 1949) the exact distribution of $D^2_{p+q}-D_p^2$, the difference between D^2's based on $(p+q)$ and p characters, was obtained under some conditions in the form of a hypergeometric series. In this paper some approximate forms of the general distribution of $D^2_{p+q}-D_p^2$ have been considecd and their use indicated.

Some computational procedures for the calculation of the D^2 statistics and discriminant functions have also been illustrated.

The following notations will be used throughout.

(1) n_1 and n_2 are the sample sizes for the first and second populations.

$$N = n_1+n_2, c = n_1n_2/(n_1+n_2).$$

(2) The difference in the mean values of the i-th character of the two populations is denoted by d_i

(3) s_{ij} is the estimate based on f degrees of freedom of the covariance between the i-th and j-th characters. If this estimate is derived from the above two samples only, then $f = n_1+n_2-2$.

(4) The statistics* defined below will have have the following standard notations.

$$D_p^2 = \Sigma\Sigma s^{ij}d_id_j$$

$$T_p = \frac{c}{f}D_p^2$$

$$U_{q\cdot p} = \frac{1+T_{p+q}}{1+T_p}-1$$

$$W_{q\cdot p} = T_{p+q}-T_p$$

The suffixes of the statistics T_p, $U_{q\cdot p}$ and $W_{q\cdot p}$ may be dropped except when several statistics with different p's and q's are considered. The population values of D_p^2

*T_p is chosen to correspond to Hotelling's T. In my earlier paper (Rao, 1949), I have denoted the statistic W by the lower case letter. The notations in capital letters seem to be convenient.

and D^2_{p+q} are denoted by

$$\Delta^2_p = \beta^2 \text{ and } \Delta^2_{p+q} = \alpha^2+\beta^2$$

so that α^2 denotes the additional distance due to q characters

2. Some Aspects of the Distribution of W

The joint distribution of $R = 1/(1+U)$ and $S = 1/(1+T)$ as obtained in (Rao, 1949, p.348) is

$$e^{-c\beta^2/2} \sum_{r=0}^{\infty} \frac{1}{r!} \left(\frac{c\beta^2}{2}\right)^r B\left(\frac{f-p+1}{2}, \frac{p}{2}+r\right) dS$$

$$\times e^{-c\alpha^2 S/2} \sum_{t=0}^{\infty} \frac{1}{t!} \left(\frac{c\alpha^2 S}{2}\right)^t B\left(\frac{f-p-q+1}{2}, t+\frac{q}{2}\right) dR$$

For testing the null hypothesis $\alpha=0$ two statistics U and W are proposed in the earlier paper. The main advantage in using U is that it provides an exact test of significance when nothing is known about β. This is due to the fact that the distribution of U when $\alpha=0$ does not contain the parameter β. The significance of an observed U can be tested by entering the quantity

$$F = \frac{f-p-q+1}{q} U$$

in the variance ratio table with q and $f-p-q+1$ drgrees of freedom. W has some theoretical advantage over U because it provides a more efficient estimate of α. The exact distribution of W is not so simple as that of U but in large samples it follows some known types and further the distribution tends to be independent of the parameter β.

2.1 *Moments of* W

The joint distribution of R and S when $\alpha = 0$ is

$$e^{-c\beta^2/2} \sum_{r=0}^{\infty} \frac{1}{r!} \left(\frac{c\beta^2}{2}\right)^r B\left(\frac{f-p+1}{2}, \frac{p}{2}+r\right) dS$$

$$\times B\left(\frac{f-p-q+1}{2}, \frac{q}{2}\right) dR$$

The statistic W is connected with R and S by the relation

$$W = (1-R)/RS = US$$

$$E(W^t) = E[(1-R)^t R^{-t} S^{-t}]$$

$$= \frac{\Gamma\left(\frac{q}{2}+t\right)}{\Gamma\left(\frac{q}{2}\right)} \cdot \frac{\Gamma\left(\frac{f-p-q+1}{2}-t\right)}{\Gamma\left(\frac{f-p-q+1}{2}\right)} \frac{\Gamma\left(\frac{f-p+1}{2}-t\right)}{\Gamma\left(\frac{f-p+1}{2}\right)} \frac{\Gamma\left(\frac{f+1}{2}\right)}{\Gamma\left(\frac{f+1}{2}-t\right)} \qquad (2.11)$$

$$\times\, e^{-c\beta^2/2}\,{}_1F_1\left(\frac{f+1}{2},\frac{f+1}{2}-t,\frac{c\beta^2}{2}\right) \qquad \ldots \quad (2.12)$$

2.2 *Large sample distribution of* W

When f is large the expression

$$e^{-c\beta^2/2}\,{}_1F_1\left(\frac{f+1}{2},\frac{f+1}{2}-t,\frac{c\beta^2}{2}\right)\to 1$$

so that $E(W^t)$ can be approximately replaced by the expression (2.11) which provides exact moments when $\beta=0$. To test the significance of the statistic W the statistic $w=W/(1+W)$ can be referred to the distribution

$$\frac{\Gamma\left(\frac{f-p+q+1}{2}\right)\Gamma\left(\frac{f+1}{2}\right)}{\Gamma\left(\frac{q}{2}\right)\Gamma\left(\frac{f+q+1}{2}\right)\Gamma\left(\frac{f-p-q+1}{2}\right)}\, w^{\frac{q}{2}}(1-w)^{\frac{f-p-q+1}{2}-1}$$

$$\times\, {}_2F_1\left(\frac{p}{2},\frac{f-p+1}{2};\frac{f+q+1}{2},w\right)dw \qquad (2.21)$$

obtained earlier in (Rao, 1949) with W in place of w.

2.2a *First approximation to* (2.21)

Using Barnes' generalization of Stirling's approximation

$$\log\Gamma(x+h)=\log\sqrt{2\pi}+(x+h-\tfrac{1}{2})\log x-x+0\left(\frac{1}{x}\right)$$

we approximate the logarithm of (2.11) to

$$\log\Gamma\left(\frac{q}{2}+t\right)-\log\Gamma\left(\frac{q}{2}\right)+t\log\{2(f+1)/(f-p-q+1)(f-p+1)\}$$

or the original expression to

$$\left\{\frac{2(f+1)}{(f-p-q+1)(f-p+1)}\right\}^t\frac{\Gamma\left(\frac{q}{2}+t\right)}{\Gamma\left(\frac{q}{2}\right)}$$

This shows that the statistic

$$\frac{(f-p-q+1)(f-p+1)}{(f+1)}\,W \qquad (2.22)$$

can be used as χ^2 on q degrees of freedom.

2.2b *Second approximation to* (2.21)

A second approximation is obtained by replacing the expression (2.11) by

$$\left(\frac{f+1}{f-p+1}\right)^t \frac{\Gamma\left(\frac{q}{2}+t\right)}{\Gamma\left(\frac{q}{2}\right)} \frac{\Gamma\left(\frac{f-p-q+1}{2}-t\right)}{\Gamma\left(\frac{f-p-q+1}{2}\right)}$$

which is the t-th moment of the variance ratio statistic

$$F = \left\{\frac{f-p+1}{f+1} W\right\} \frac{f-p-q+1}{q} \tag{2.23}$$

based on q and $f-p-q+1$ degrees of greedom. The approximation (2.23) is slightly better than (2.22) when q is not small.

2.3 *The exact distribution of* W

The r-th term in the expansion of $E(W^t)$ obtained in section 2.1 is the product of

$$\frac{1}{r!} e^{-c\beta^2/2} \left(\frac{c\beta^2}{2}\right)^r \tag{2.31}$$

and

$$\frac{\Gamma\left(\frac{q}{2}+t\right)}{\Gamma\left(\frac{q}{2}\right)} \frac{\Gamma\left(\frac{f-p-q+1}{2}-t\right)}{\Gamma\left(\frac{f-p-q+1}{2}\right)} \frac{\Gamma\left(\frac{f-p+1}{2}-t\right)}{\Gamma\left(\frac{f-p+1}{2}\right)} \frac{\Gamma\left(\frac{f+1}{2}+r\right)}{\Gamma\left(\frac{f+1}{2}+r-t\right)} \tag{2.32}$$

This is the moment function of $W=w/(1-w)$ where w has the distribution

$$\frac{\Gamma\left(\frac{f-p+q+1}{2}\right)}{\Gamma\left(\frac{q}{2}\right)\Gamma\left(\frac{f-p-q+1}{2}\right)} w^{\frac{q}{2}-1} (1-w)^{\frac{f-p-q+1}{2}-1} \quad \dots \tag{2.33}$$

$$\times e^{-c\beta^2/2} \sum_{s=0}^{\infty} \frac{1}{s!}\left(\frac{c\beta^2}{2}\right) \frac{\Gamma\left(\frac{f+1}{2}+s\right)}{\Gamma\left(\frac{f+q+1}{2}+s\right)} {}_2F_1\left(\frac{p}{2}+s, \frac{f-p+1}{2}; \frac{f+q+1}{2}, w\right) dw$$

2.4 *Approximate distribution of W involving* β

Using Barnes' approximation the expression (2.32) can be reduced to

$$\left\{\frac{2(f+1+2r)}{(f-p-q+1)(f-p+1)}\right\}^t \frac{\Gamma\left(\frac{q}{2}+t\right)}{\Gamma\left(\frac{q}{2}\right)}$$

which is the t-th moment of the distribution

$$\left(\frac{a_r}{2}\right)^{\frac{q}{2}} \frac{1}{\Gamma\left(\frac{q}{2}\right)} e^{-a_r W/2} W^{\frac{q}{2}-1} dW \qquad \ldots \quad (2.41)$$

where $$a_r = \frac{(f-p-q+1)(f-p+1)}{(f+1+2r)}$$

Combining the expression (2.31) with (2.41) the distribution of W is obtained in the form of the infinite series

$$\frac{e^{-c\beta^2/2}}{2^q\,\Gamma\left(\frac{q}{2}\right)} W^{\frac{q}{2}-1} \sum_{r=0}^{\infty} \frac{1}{r!}\left(\frac{c\beta^2}{2}\right)^r a_r^q\, e^{-a_r W/2}\, dW \qquad \ldots \quad (2.42)$$

This is a series involving χ^2 distributions with the dominant term

$$\frac{a_0^q}{2^q\Gamma\left(\frac{q}{2}\right)} e^{-a_0 W/2}\, W^{\frac{q}{2}-1} dW$$

which corresponds to the approximation obtained in (2.22). It is easy to see that for any given W the probability of exceeding that value according to (2.22) is always smaller than that according to (2.42). This means that that the approximation (2.22) overestimates significance. The magnitude of this overestimation will be greater for higher values of β as should be expected. Similarly it can be shown that the second approximation obtained in (2.23) also overestimates significance when $\beta \neq 0$. If the samples are large enough the effect of non-zero β will be small.

3. *Transformation of correlated variables*

In appendix 5 in (Mahalanobis, Majumdar and Rao, 1949) the author has given a simple method of constructing a set of uncorrelated variables from a mutually correlated set. If $x_1, x_2, \ldots$ represent the correlated variables the transformation suggested is of the type

$$\begin{aligned} Y_1 &= x_1 \\ Y_2 &= x_2 - a_{21}Y_1 \\ Y_3 &= x_3 - a_{32}Y_2 - a_{31}Y_1 \end{aligned} \qquad \ldots \quad (3.1)$$

The variables $Y_1, Y_2 \ldots$ are all uncorrelated and they are constructed successively one after the other using the variances and covariances of the x's. The coefficients $a_{i1}, \ldots, a_{ii-1}$ in Y_i are also successively calculated one after the other, any coefficient

a_{ij} depending on all a_{rs}, $r<i$ and $s<j$. Using the transformation (3.1) one obtains the Y values for any given set of x values by successive substitutions. It is, however, not possible to compute any Y_i directly from the x's without evaluating Y_r for $r<i$. Wold (1950) has shown that by a series of recursive calculations the expressions for Y's in terms of x's can be obtained from the transformation (3.1). By this method, the expressions for Y_1, Y_2,... are obtained successively one after the other. They can be simultaneously evaluated by following the simple devise of sweep out as illustrated below.

Consider the first five equations given on page 152 Sankhya, Vol. 9, pts. 2 & 3, 1949. Rearranging the Y's and representing the standardised characters *hl*, *hb*, *b,b*, *nh* and *nb* by x_1, x_2, x_3, x_4 and x_5 respectively these equations may be written in the form of a matrix.

The matrix of coefficients

Y_1	Y_2	Y_3	Y_4	Y_5	=	x_1	x_2	x_3	x_4	x_5
1	.	.	.	.		1	.	.	.	.
.1982	1	.	.	.		.	1	.	.	.
.2702	.5052	1	.	.		.	.	1	.	.
.1758	.1443	.0976	1	.		.	.	.	1	.
.1930	.1073	.2467	−.0232	1		.	.	.	.	1 .

On sweeping out the first five columns which can be easily done because the matrix on the left is semi-diagonal with diagonal elements all unity we obtain the resulting matrix

Y_1	Y_2	Y_3	Y_4	Y_5	=	x_1	x_2	x_3	x_4	x_5
1	.	.	.	.		1	.	.	.	.
.	1	.	.	.		−.1982	1	.	.	.
.	.	1	.	.		−.1791	−.5052	1	.	.
.	.	.	1	.		−.1297	−.0950	−.0975	1	.
.	.	.	.	1		−.1305	.0150	−.2490	.0232	1

which gives the expressions for Y's in terms of x's

An alternative method which directly yields the functions of x's is suggested by the following theoretical considerations. Let the dispersion (variance covariance) matrix of the variates $x_1, x_2, \ldots x_p$ be denoted by

$$\Lambda = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1p} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ \lambda_{p1} & \lambda_{p2} & \cdots & \lambda_{pp} \end{pmatrix}$$

Consider the extended matrix

$$\begin{matrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1p} & x_1 \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2p} & x_2 \\ \cdots & \cdots & \cdots & \cdots. & \cdots \\ \lambda_{p1} & \lambda_{p2} & \cdots & \lambda_{pp} & x_p \end{matrix}$$

Taking λ_{11} as the first pivotal element replace the first row by

$$\begin{matrix} 1 & \frac{\lambda_{12}}{\lambda_{11}} & \cdots & \frac{\lambda_{1p}}{\lambda_{11}} & \frac{x_1}{\lambda_{11}} \end{matrix}$$

Sweeping out the first column using the first pivotal row we obtain the reduced matrix

$$\begin{matrix} \lambda'_{22} & \cdots & \lambda'_{2p} & x_2' \\ \cdots & \cdots & \cdots & \cdots \\ \lambda'_{p2} & \cdots & \lambda'_{pp} & x_p' \end{matrix}$$

where
$$\lambda'_{ij} = \lambda_{ij}-\frac{\lambda_{i1}}{\lambda_{11}}\lambda_{1j},\ x_i' = x_i-\frac{\lambda_{i1}}{\lambda_{11}}x_1$$

Now
$$V(x_i') = V(x_i)-2\frac{\lambda_{i1}}{\lambda_{11}}\text{Cov}\,(x_ix_1)+\left(\frac{\lambda_{i1}}{\lambda_{11}}\right)^2 V(x_1)$$

$$= \lambda_{ii}-\frac{\lambda_{i1}^2}{\lambda_{11}} = \lambda_{ii}'$$

Similarily $\quad \text{Cov}(x_ix_j) = \lambda_{ij}'$

This shows that the reduced matrix at any stage is the dispersion matrix of the new variables on the right hand side provided the first matrix is the dispersion matrix of the original variables. This property has been discussed by the author (Rao, 1945) in connexion with solution of normal equations and their intrisic properties. Also

$$\text{Cov}(x_1x_i') = \text{Cov}(x_1x_i)-\frac{\lambda_{i1}}{\lambda_{11}}V(x_1)$$
$$= \lambda_{i1}-\lambda_{i1} = 0$$

so that the new variables are all uncorrelated with the variable of the pivotal row. We now consider the second pivotal row

$$\begin{matrix} 1 & \frac{\lambda_{23}'}{\lambda_{22}'} & \cdots & \frac{\lambda_{2p}'}{\lambda_{22}'} & \frac{x_2'}{\lambda_{22}'} \end{matrix}$$

and obtain the further reduced matrix

$$\begin{matrix} \lambda_{33}'' & \cdots & \lambda_{3p}'' & x_3'' \\ \cdots & \cdots & \cdots & \cdots \\ \lambda_{p3}'' & \cdots & \lambda_{pp}'' & x_p'' \end{matrix}$$

We thus obtain the variables

$$x_1,\ x_2',\ x_3'',\ldots\ldots$$

with variances

$$\lambda_{11}, \lambda_{22}' \lambda_{33}''......$$

They are all mutually uncorrelated as shown above and and further x_2' depends on x_1 and x_2 and x''_3 on x_1, x_2 and x_3 only, and so on. Thus the transformation is of the type considered earlier. The method of construction is illustrated below. The correlation* matrix of the variables x_1, x_2, x_3, x_4, x_5 considered before is given in Table 1 with an extended unit matrix.

The function for $Y_1, Y_2, ...$ obtained in Table 1 are same as those derived earlier to the order of significant figures retained in the computations. We thus obtain a relatively simple scheme for obtaining uncorrelated linear functions of the original variables provided the number and variables to be included are fixed in advance. In the earlier method the transformed variables are calculated one after the other so that we are free to choose the variable to be added at any stage and in any order we like. There are a few problems where the decision to add a new character depends on tests to be made with help of the transformed variates up to that stage. In situations like this only the earlier method is open to us. It is enough to compute the transformation (3.1) in such a case since successive values of $Y_1, Y_2, ...$ will be obtained. There is no need to express the Y's as functions of x only. In problems where a transformation of a chosen set of correlated variables is required the method of Table 1 is the best.

Having obtained $Y_1, Y_2, ..., Y_5$ directly as functions of the original variables if we want to extend the transformation to a sixth variable x_6 then we write

$$Y_6 = x_6 - a_{65}Y_5 - a_{64}Y_4 - a_{63}Y_3 - {}_{62}Y_2 - a_{61}Y_1$$

as in the earlier method. The coefficients are determined from the equations

$$\text{Cov}(x_6 Y_i) = a_{6i} V(Y_i)$$

Since Y_i is a known function of the x's it is easy to calculate Cov $(x_6 Y_i)$ and $V(Y_i)$ is directly available from Table 1. Let the new variable x_6 have the correlations .1537, .1308, .1575, .2910 and .1139 with x_1, x_2, x_3, x_4, and x_5 respectively.
Then

$$\text{Cov}\ (x_6 Y_1) = \text{Cov}(x_6 x_1) = .1537$$

$$V(Y_1) = 1,\ a_{61} = .1537$$

$$\text{Cov}\ (x_6 Y_2) = \text{Cov}\, x_6(x_2 - .1982x_1) = .1003$$

$$V(Y_2) = .9607,\ a_{62} = .1044$$

$$\text{Cov}\ (x_6 Y_3) = \text{Cov}\ x_6(x_3 - .5053x_2 - 1791x_1) = .0639$$

$$V(Y_3) = .0707,\ a_{63} = .0944$$

$$\text{Cov}\ (x_6 Y_4) = \text{Cov}\, x_6(x_4 - .0975x_3 - .0951x_2 - .1297x_1) = .2433$$

$$V(Y_4) = .9427,\ a_{64} = .2581$$

*The correlation matrix is being considered because the variables $x_1, x_2, ..., x_5$ have been already standardised. Otherwise the variance-covariance matrix should be taken.

Table 1. Pivotal condensation method for the construction of a transformation to an uncorrelated set

Row	Correlation matrix					x_1	x_2	x_3	x_4	x_5	check
01	1	.1982	.2972	.1758	.1930	1				= Y_1	2.8462
02		1	.5407	.1735	.1413		1				3.0537
03			1	.1852	.2729			1			3.2780
04				1	.0438				1		2.5783
05					1					1	2.6510
10	1	.1982	.2792	.1758	.1930	1					2.8402
11		.9607	.4854	.1387	.1030	−.1982	1			= Y_2	2.4896
12			.9220	.1301	.2190	−.2792		1			2.4833
13				.9691	.0009	−.1758			1		2.0779
14					.9628	−.1930				1	2.1017
20		1	.5053	.1444	.1072	−.2063	1.0409				2.5914
21			.6767	.0660	.1670	−.1701	−.5053	1		= Y_3	1.2254
22				.9491	−.0050	−.1472	−.1444		1		1.7185
23					.9518	−.1718	−.1072			1	1.8348
30			1	.0975	.2468	−.2647	−.7407	1.4778			1.8108
31				.9427	−.0213	−.1297	−.0951	−.0975		= Y_4	1.5990
32					.9106	−.1276	.0175	−.2468		1	1.5324
40				1	−.0226	−.1376	−.1009	−.1034	1.0008		1.6902
41					.9101	−.1305	.0154	−.2499	.0226	1 = Y_5	1.5685

1. The elements below the diagonal are omitted because the matrices at all stages are symmetrical.

2. The rows 10, 20, 30, 40 represent the pivotal rows at each stage of reduction. Following each pivotal row is given the reduced matrix. The first row in the reduced matrix supplies a linear function of the variables whose variance is the pivotal element (underlined) in that row. Thus Y_1, Y_2, Y_3, Y_4, and Y_5 have 1.0000, .9607 .6767, 9427 and .0101 as their variances

3. The computations can be compactly represented by ommitting the matrix for functions of original variables and accomodating the figures below the diagonal in the left hand side matrix as shown in Table 2. This can be done only when one has acquired sufficient experience in computations of this nature.

4. The dispersion matrix should be used instead of the correlation matrix if the original variables are not already standardized.

$$\text{Cov}\,(x_6 y_5) = \text{Cov}\; x_6(x_5 + .0226x_4 - .2490x_3 + .0154x_2 - .1305x_1) = .0632$$

$$V(Y_5) = .9101, \quad a_{65} = .0694$$

$$Y_6 = x_6 - .0694Y_5 - .2581Y_4 - .0944Y_3 - .1044Y_2 - .1537Y_1$$

$$V(Y_6) = 1 - \Sigma a_{6i}\; \text{Cov}\,(x_6 x_i) = 1 - .1073 = .8927$$

In the earlier paper (Rao, 1949) is given a method of pivotal condensation by which independent contributions to D^2 with the successive addition of characters can be evaluated in a simple manner. It is also of importance to find out the discriminant function at each stage. This is possible with the help ot a slightly modified computational scheme as given by Aitken (1933). Choosing the illustrative example given in (Rao, 1949) the scheme of computation is presented in Table 2.

In this table the matrix in between two pivotal rows is the reduced matrix using the pivotal row above it. The last row at each stage of reduction supplies the D^2 value and the corresponding discriminant function. These are collected below in Table 3 with a further check column.

Table 3. Values of D^2 and discriminant function coefficients.

Row	Successive values of D^2	Discriminant function $L(x)$	Value of $L(x)$ when x's are replaced by difference in means
14	4.4286	$4.7619x_1$	4.4286
23	76.7082	$-11.1013x_1 + 31.1052x_2$	76.7081
32	92.3808	$-4.4082x_1 + 31.1393x_2$ $-14.2141x_3$	92.3810
41	103.2119	$-3.0092x_1 + 21.7641x_2$ $-18.0006x_3 + 30.8573x_4$	103.2119

The second and fourth columns agree thus providing a final check on all the calculations.

A note on the method of pivotal condensation

The method of reducing a matrix by the method of pivotal condensation seems to yield many interesting results. Since the various steps in the process can be calculated is a routine manner it commends itself as the ideal computational technique. It is worthwhile exploring the various ways of using this technique in statistical computations.

The simplest use of the method of pivotal condensation is in the evaluation of the value of the determinant which is equal to the product of all the pivotal elements.

In a number of multivariate computations the inverse of a matrix is needed. This can be easily done by appending a unit matrix to the original one and reducing the latter by the method of pivotal condensation, each time 'sweeping out' all the other elements in the column i.e., all those above and below the pivotal element.

TABLE 2. PIVOTAL CONDENSATION METHOD FOR SUCCESSIVE D²'s AND DISCRIMINANT FUNCTIONS L(x)

Row		Dispersion matrix x_1	x_2	x_3	x_4	Difference in means	Sum including the indented	Check excluding the indented.
01		.1953	.0996	.0922	.0331	.930	1.350200	
02			.1255	.0472	.0396	2.798	3.109900	
03				.1211	.0252	−.658	−.372300	
04					.0251	1.083	1.203000	
05						0.000	3.150000	
10		1	.509985	.472094	.169483	4.761904		6.913406
11		.509985	.074705	.000179	.022719	2.323714	2.931302	2.421317
12		.472094		.077573	.009574	−1.097047	.537627	−1.009721
13		.169483			.019490	.922381	1.143647	.974164
14	L_1 =	4.761904				−4.428571 = $-D_1^2$	−2.482381	2.279523
20		6.820638	1	.002396	.304116	31.105200		39.238350
21		.470872	.002396	.077572	.000519	−1.102615	−.542256	−.544052
22		.014389	.304116		.012581	.215702	.550307	.252191
23	L_2 =	−11.101250	31.105200			−76.708160 = $-D_2^2$	−57.609123	−88.714323
30		6.070128	.030887	1	.122712	−14.214085		−6.990358
31		− .043393	.303822	.122712	.011413	.351006	.745560	.622848
32	L_3 =	−4.408236	31.130256	−14.214085		−92.380813 = $-D_3^2$	−70.512872	−65.298787
40		−3.802068	26.620696	10.751950	1	30.754929		65.325507
41	L_4 =	−3.069247	21.764139	−18.000042	30.754029	−103.211900 = $-D_4^2$		102.517650

1. The rows 10, 20, 30, and 40 are the pivotal rows at each stage

2. After swooping out the first column fill the column by the elements in the 1st pivotal row. These are indented as shown above. Retain these elements in swooping out the second column at the second stage. In the reduced matrix fill in the second column by elements in the second pivotal row. Retain them in the sweep out processes at subsequent stages.

3. The sums in the last but one column are used in obtaining the elements of the check column at each stage of reduction. Thus the reduced values of 2.931302, .537627,...... in rows 11, 12, ... are written in the check column in rows 20, 21, ..., c.f., 39.238350, —0.544052, ... From these values the previous column is built up by adding the indented column.

A very important use of this method is in the successive evaluation of the regression equations of a variable (y) on x_1; x_1 and x_2; x_1, x_2, and x_3; and so on. This method is suggested by Aitken. Since the discriminant function can be regarded as a regression equation the same method could be used. A slight modification has been made to save space and minimise the number of entries (see Table 2). With some practice this method can be conveniently carried out.

A new use of the method of pivotal condensation is found in Table 1 of this article. The reduction of a dispersion matrix with an appended unit matrix seems to yield a semi-diagonal transformation of the original variables to an uncorrelated set.

REFERENCES

AITKEN, A. C. (1933). On fitting polynomials to weighted data by least squares. *Proc. Roy. Soc. Edin.*, 54, 1

MAHALANOBIS, P. C., MAJUMDAR, D. N., and RAO, C. R. (1949). Anthropometric survey of the United Provinces 1941: A statistical study. *Sankhya*, 9, 90.

RAO, C. R. (1949). On some problems arising out of discrimination with multiple characters. *Sankhya*, 9, 343.

WOLD, H. (1950). A large sample test for moving averages, *J. Roy. Stat. Soc* (in Press) ·