ON SOME PROBLEMS ARISING OUT OF DISCRIMINATION WITH MULTIPLE CHARACTERS

By C. RADHAKRISHNA RAO
Statistical Laboratory, Calcutta

1. Introduction

The methods of multivariate analysis have been found extremely useful in a variety of problems arising in biological, psychological and economic research. Some special applications of these methods have been considered in detail by Bartlett (1947, 1948), Fisher (1936), Geary (1948), Rao (1948a, 1948b), Rao and Slater (1949) and Tintner (1946). Recently, in an article by Mahalanobis, Majumdar and Rao (1949), use has been made of Mahalanobis' distance, known as the D^2-statistic, in arriving at a reasonable classification of 22 inbreeding groups of men living in the United Provinces of India. In this investigation, there were about 12 characters and it was found that the general classification obtained with about 8 or 9 characters remained practically unaffected when more characters were used in the analysis. Perhaps the number of characters which can give stable configurations of the groups could be further reduced by a suitable choice of the characters. This is due to the fact that the increase in D^2 due to the addition of more characters to a suitably chosen panel was not appreciable except in a few cases. The increases in D^2 when the number of characters is increased from 9 to 12 are given in Table 12.2 of the article by Mahalanobis, Majumdar and Rao (1949).

Unless D^2 converges to some fixed value with increase in the number of characters no stable classification of a given set of groups can be obtained by using a relatively smaller number of characters. This cannot be mathematically studied but can be emperically verified in any situation. This aspect of D^2 as a necessary condition for obtaining valid classifications of groups was stated in an axiomatic form by Mahalanobis (1937).

I have shown, elsewhere (Rao, 1948b), that the frequency of misclassification by using the discriminant function between two populations is related to D^2 between them in a simple manner. It is seen that if the increase in D^2 is not appreciable, the reduction in errors of classification by using more characters is negligible.

Thus, it appears in problems of classification (*i.e.* of specifying an individual as a member of one of many groups to which he can possibly belong or classifying the groups themselves into some significant system based on the configuration of the various characteristics) the information gained by increasing the number of characters after a certain stage is not appreciable. The extra information, thus gained, may

not be commensurate with the cost of obtaining the measurements on the additional characters and the computational labour involved in utilizing them.

It is, however, of some interest to examine how far the tests of significance are affected by increase in the number of characters. It does not seem to be, always, the more the better as shown in the next section.

2. AN EXAMPLE OF DISCRIMINATION WITH THE LENGTHS OF FEMUR AND HUMERUS

Table 1 gives the mean values of lengths of Femur and Humerus of 20 Indian and 27 Anglo-Indian skeletons.

TABLE 1. MEAN VALUES OF FEMUR AND HUMERUS LENGTHS

	sample size	mean length of	
		Femur	Humerus
Anglo-Indians	27	460.4	335.1
Indians	20	444.3	323.2
difference		16.1	11.9

The pooled estimates (on 45 degrees of freedom) of standard deviations of Femur and Humerus lengths are 23.7 and 18.2 respectively and the correlation 0.8675.

To test for the difference in Femur lengths of the two populations the F statistic with 1 and 45 degrees of freedom is

$$F = \frac{27\times 20}{27+20}\frac{(16.1)^2}{(23.7)^2} = 5.301$$

which is significant at the 5% level. The corresponding statistic for testing the difference in Humerus length is

$$F = \frac{27\times 20}{27+20}\frac{(11.9)^2}{(18.2)^2} = 4.901$$

which is also significant at the 5% level

To test jointly for the differences in the mean lengths of Femur and Humerus, Hotelling's T or Mahalanobis' D^2 has to be used. The statistic

$$\frac{47-3}{47-2}\cdot\frac{27\times 20}{27+20}\cdot\frac{1}{2(1-\rho^2)}\left\{\left(\frac{16.1}{23.7}\right)^2 - 2\rho\left(\frac{16.1}{23.7}\right)\left(\frac{11.9}{18.2}\right) + \left(\frac{11.9}{18.2}\right)^2\right\}$$

$$= 2.685 \text{ (for } \rho = .8675)$$

can be used as a variance ratio for 2 and 44 degrees of freedom for which the 5% significant value is slightly above 3.21. The quantity calculated above is not significant showing thereby that there is no evidence in the data to show that the populations are different!

Here is a dangerous situation where the inclusion of an extra character is not beneficial in discriminating between the two populations. It appears that some care

is needed in the choice of characters for tests of significance also. A general investigation of this problem is undertaken in the rest of the paper.

3. The distribution of D^2 and allied statistics

The following symbols will be used throughout.

(1) n_1, n_2 are the sample sizes for the first and second populations. $N = n_1 + n_2$, $c = n_1 n_2/(n_1 + n_2)$.

(2) $\bar{x}_{i1}$, $\bar{x}_{i2}$ are the mean values of the i-th character for the first and second populations. The difference $\bar{x}_{i1} - \bar{x}_{i2}$ is denoted by l_i.

(3) S_{ij} is the pooled corrected sum of products within the groups for the variates x_i and x_j. The estimated covariance $S_{ij}/(n_1 + n_2 - 2)$ is denoted by s_{ij}.

(4) The D^2 statistic for any given p variates is defined by

$$D^2_p = \sum_{i,j=1}^{p} s^{ij} d_i d_j$$

where (s^{ij}) is the $p \times p$ matrix reciprocal to (s_{ij}). The population value of D^2_p is denoted by

$$\Delta^2_p = \sum_{i,j=1}^{p} \sigma^{ij} \delta_i \delta_j$$

where (σ_{ij}) is the population value of (s_{ij}) and δ_i that of d_i. For some computational aspects of the D^2 statistic the reader is referred to Appendix 1 at the end of the paper.

Consider the following problem of testing the difference in a $(p+1)$th character between two populations after eliminating the observed differences in some p characters in samples of sizes n_1 and n_2 from two populations. This is a problem in the theory of least squares, the general technique of which is given by the author in (Rao, 1946a). We set up the observational equations

$$E(x_{p+1}) = \alpha_1 + \beta_1 x_1 + \ldots + \beta_p x_p$$

and

$$E(x_{p+1}) = \alpha_2 + \beta_1 x_1 + \ldots + \beta_p x_p$$

for observations from the first and second populations respectively. The hypothesis to be tested is $\alpha_1 = \alpha_2$. The residual sum of squares with $n_1 + n_2 - 2 - p$ degrees of freedom is seen to be

$$R_p = \frac{|S_{ij}|_{p+1}}{|S_{ij}|_p}, \quad \text{where } |S_{ij}|_q = \begin{vmatrix} S_{11} \ldots S_{1q} \\ \ldots \\ S_{q1} \ldots S_{qq} \end{vmatrix}$$

If the hypothesis is true then the residual sum of squares with $n_1 + n_2 - 1 - p$ degrees of freedom is

$$R'_p = \frac{|S_{ij} + c d_i d_j|_{p+1}}{|S_{ij} + c d_i d_j|_p}, \quad \text{where } c = \frac{n_1 n_2}{n_1 + n_2}$$

The estimate of $\alpha_1-\alpha_2$ is

$$z = d_{p+1} - b_1 d_1 - \ldots - b_p d_p$$

where $b_1, \ldots, b_p$ are the estimates of $\beta_1, \ldots, \beta_p$. This has the variance

$$\frac{1}{c}\,\frac{|\,S_{ij}+cd_id_j\,|_p}{|\,S_{ij}\,|_p}\;\sigma^2$$

where σ^2 is the variance of x_{p+1} for a given set of $x_1, \ldots, x_p$. If we define

$$M_p = \frac{|\,S_{ij}+cd_id_j\,|_p}{|\,S_{ij}\,|_p} = 1+\frac{c}{n_1+n_2-2}\,D^2_p$$

then it follows by the theorem of least squares

$$R'_p = R_p + cz^2/M_p$$

Hence

$$\frac{R_p}{R'_p} = \frac{R_p}{R_p+cz^2/M_p} = \frac{M_p}{M_{p+1}}$$

R_p is distributed as χ^2 with (n_1+n_2-p-2) degrees of freedom and z is a normal variate distributed independently of R_p. If $\sigma=1$, the joint distribution of z and R_p is

$$\text{Const. } e^{-\frac{c}{2}\frac{(z-\alpha)^2}{M_p}-\frac{R_p}{2}}\,R_p^{\frac{n_1+n_2-p-4}{2}}\,dR_p\,dz$$

where $\alpha_1-\alpha_2=\alpha$. The distribution of

$$y_{p+1} = \frac{M_p}{M_{p+1}} = \frac{R_p}{R_p+cz^2/M_p}$$

obtained after making the polar transformation $\sqrt{c/M_p}\,z=r\cos\theta$ and $\sqrt{R_p}=r\sin\theta$ and integrating over r is

$$\text{Const. } e^{-c\alpha^2/2M_p}\,y_{p+1}^{\frac{N-p-1}{2}}\,(1-y_{p+1})^{-\frac{1}{2}}\,{}_2F_1\left[\frac{N-p-1}{2},\ \frac{1}{2},\ \frac{c\alpha^2}{2M_p}\,(1-y_{p+1})\right]dy_{p+1}$$

Denoting*

$$B(a,b)\,dy = y^{a-1}(1-y)^{b-1}/\beta(a,b)$$

the above distribution can be written

$$e^{-c\alpha^2/2M_p}\sum\frac{1}{t!}\left(\frac{c\alpha^2}{2M_p}\right)^t B\left(\frac{N-p-2}{2},\,t+\frac{1}{2}\right)dy_{p+1}$$

* The variable is omitted in the representation of the function by B (a, b). The variable in it is taken to be same as that in the differential element following it.

When $\alpha=0$, the distribution is simply

$$B\left(\frac{N-p-2}{2}, \frac{1}{2}\right) dy_{p+1}$$

The distributions of statistics defined below are all derivable with the help of the distribution established above using the univariate theory of least squares.

To test whether the addition of q more characters to a set of p characters increases the distance between two populations the statistic used is the ratio M_{p+q}/M_p which gives a comparison of the D^2's based on p and $p+q$ characters. The distribution of this statistic in the null case was derived in (Rao, 1946b). In this section we shall obtain the distribution assuming that there is an increase in the distance.

Let us transform the original variates $x_1, \ldots, x_p, x_{p+1}, \ldots, x_{p+q}$ to an uncorrelated set $u_1, \ldots, u_p, u_{p+1}, \ldots, u_{p+q}$ with unit variances such that $u_1, \ldots u_p$ depend only on $x_1, \ldots, x_p$. Since M_p and M_{p+q} are invariant under a linear transformation of the varintes we can derive the distribution of M_p/M_{p+q} considering the u's as variables. Without loss of generality we can take the probability densities in the two populations to be

$$\text{Const. } e^{-\frac{1}{2}[(u_1-\beta)^2+u_2^2+\ldots+u_p^2+(u_{p+1}-\alpha)^2+u_{p+2}^2\ldots+u_{p+q}^2]}$$

and

$$\text{Const. } e^{-\frac{1}{2}[u_1^2+\ldots+u_p^2+u_{p+1}^2+\ldots+u_{p+q}^2]}.$$

The quantities α and β are defined by

$$\beta^2=\Delta_p^2,\ \alpha^2+\beta^2=\Delta_{p+q}^2$$

so that α^2 is the addition to the square of the distance. It may be noted that if the whole distance between two populations is accounted by the first p characters alone any linear function of characters uncorrelated with the first p characters has the same mean value for both the populations. In such a situation the parameter α, standing for the additional distance, is zero. This is the hypothesis assumed in my earlier papers (Rao, 1946b & 1948a).

First let us consider $u_1, \ldots, u_p$ as fixed. Then by the use of the distribution established above we obtain the joint distribution of $y_{p+q}, \ldots, y_{p+1}$ as

$$B\left(\frac{N-p-q-1}{2}, \frac{1}{2}\right) dy_{p+q}\ B\left(\frac{N-p-q}{2}, \frac{1}{2}\right) dy_{p+q-1} \ldots B\left(\frac{N-p-3}{2}, \frac{1}{2}\right) dy_{p+2}$$

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

Considering term by term in the infinite series above and applying the result of Appendix 3, we obtain the distribution of the ratio

$$R = \frac{M_p}{M_{p+q}} = y_{p+q}\, y_{p+q-1} \cdots y_{p+1}$$

as

$$e^{-c\alpha^2/2M_p} \sum_{t=0}^{\infty} \frac{1}{t!} \left(\frac{c\alpha^2}{2M_p}\right)^t B\left(\frac{N-p-q-1}{2},\ t+\frac{q}{2}\right) dR \qquad (3.1)$$

This is the conditional distribution of R given $1/M_p$.

We now observe that the distribution of $S = 1/M_p = y_1, \ldots, y_p$ is derivable in an exactly similar manner from the joint distribution of $u_1, u_2, \ldots, u_p$. This is

$$e^{-c\beta^2/2} \sum_{r=0}^{\infty} \frac{1}{r!} \left(\frac{c\beta^2}{2}\right)^r B\left(\frac{N-p-1}{2},\ r+\frac{p}{2}\right) dS \qquad (3.2)$$

The above method appears to be a simple way of obtaining the non-null distribution of D^2 or its related function S which was derived earlier by Bose and Roy (1938) and Hsu (1938). The joint distribution of R and S which now replaces $1/M_p$ in the conditional distribution (3.1) is

$$\begin{aligned} &e^{-c\beta^2/2} \sum_{r=0}^{\infty} \frac{1}{r!} \left(\frac{c\beta^2}{2}\right)^r B\left(\frac{N-p-1}{2},\ r+\frac{p}{2}\right) dS \\ &\times e^{-c\alpha^2 S/2} \sum_{t=0}^{\infty} \frac{1}{t!} \left(\frac{c\alpha^2 S}{2}\right)^t B\left(\frac{N-p-q-1}{2},\ t+\frac{q}{2}\right) dR \end{aligned} \qquad (3.3)$$

The distribution of R in the case $\alpha = 0$, is independent of S as observed in an earlier paper (Rao, 1946b). To obtain the unconditional distribution of R the above expression has to be integrated for S from 0 to 1,

Defining

$$\phi(u) = \int_0^1 \exp[-uS - c\beta^2/2] \left\{ \sum_{r=0}^{\infty} \frac{1}{r!} \left(\frac{c\beta^2}{2}\right)^r B\left(\frac{N-p-1}{2}, \frac{p}{2}+r\right) \right\} dS$$

and

$$\phi^{(r)}(u) = \frac{d^r\phi}{du^r}$$

The distribution of R can be written as

$$\sum_{t=0}^{\infty} \frac{(-1)^t}{t!} \phi^{(t)}\left(\frac{c\alpha^2}{2}\right) B\left(\frac{N-p-q-1}{2},\ t+\frac{q}{2}\right) dR \qquad (3.4)$$

Expanding $e^{-c\alpha^2 S/2}$ in powers of S and integrating the expressions term by term in (3.3), the distribution in (3.4) can be thrown into the alternate form*

$$\sum_{r=0}^{\infty}\sum_{t=0}^{\infty}\frac{(-1)^r}{r!}\frac{1}{t!}\left(\frac{c\alpha^2}{2}\right)^{r+t}\frac{\Gamma\left(\frac{N-p-1}{2}+r+t\right)\Gamma\left(\frac{N-1}{2}\right)}{\Gamma\left(\frac{N-p-1}{2}\right)\Gamma\left(\frac{N-1}{2}+r+t\right)}$$

$$\times {}_1F_1\left(r+t,\frac{N-1}{2}+r+t,\frac{-c\beta^2}{2}\right)B\left(\frac{N-p-q-1}{2},t+\frac{q}{2}\right)dR \quad (3.5)$$

If the dispersion matrix is estimated on f degrees of freedom while the means are based on samples of sizes n_1 and n_2 the distribution of R is obtained by replacing N by $f+2$ in the above distributions.

The problem of inference concerning α^2 : The unconditional distribution (3.4) of R contains the *nuisance parameter* β^2 so that using this distribution an exact inference on α^2 independent of β^2 is not possible. The conditional distribution of R given S is, however, independent of β^2. It is also seen that the distribution (3.2) of S is independent of α^2 so that the statistic S by itself cannot supply any information on α. Such a statistic whose distribution is independent of the parameter under consideration but which jointly with others summarise the information on the parameter is called an ancillary statistic. The existence of the ancillary statistic S and the independence of the conditional distribution of R from the nuisance parameter β^2 allow an exact conditional inference to be made on α^2. We need use only the conditional distribution of R

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

for testing any hypothesis concerning α^2 or finding its fiducial distribution. Extensive tables of the probability integral of the above distribution are under preparation in the Indian Statistical Institute at Calcutta. This will be useful in tests of significance and also in determining fiducial limits.

One point to be noted is that for testing the null hypothesis $\alpha = 0$, the conditional power of the test increases with increase in $c\alpha^2 S$ or, for a given c and α^2, with

* *Added in proof* :—When this paper was in press, R. D. Narain announced this non-null distribution, in a different form, without proof in *Current Science*, Vol. 18, 243.

increase in S, the value of the ancillary statistic. S will be a maximum if M_p is a minimum or the value of D^2_p calculated from the sample is zero. The smaller the value of D^2_p, the higher will be the power of the test for judging the significance of the additional information supplied by q more characters.

In a problem of studying the differences introduced by training, Cochran and Bliss (1948) stated that the initial I.Q.'s might be used as concomitant variates in discriminating with the help of measurements made at the end of training. They suggest that after measuring the initial I.Q.'s, a sample of students may be divided *at random* into two groups, each of which subsequently receives a different type of training. The power of discrimination with the help of final measurements will be greater if the two groups are chosen such that the mean values of the initial I.Q's in the groups are *as near as possible* so that the D^2 with the initial characters is small.

The distribution of a statistic alternative to R : As mentioned earlier, the conditional distribution of the ratio R is useful in testing the hypothesis $\alpha=0$, when β is *unknown*. When $\beta=0$, arising in some situations already mentioned, it was observed that conditional inference can be made with the help of the R-statistic although it may not be the best possible test in Neyman's sense. It may be recalled that the R-statistic is the ratio M_p/M_{p+q} which gives a comparison of the D^2-statistics based on p and $(p+q)$ characters. An alternative statistic* is

$$w=M_{p+q}-M_p \tag{3.6}$$

which measures the difference between the D^2-statistics based on p and $(p+q)$ characters. It is of interest to examine whether the statistic w is more useful than R in the particular situation $\beta=0$. Both R and w being functions of M_p and M_{p+q}, the conditional tests based on them for given S or M_p should be indentical. The unconditional distribution of R, when the null hypothesis ($\alpha=0$) is true is the same as that of the conditional distribution for given S even when β is unknown as shown in (Rao, 1946b).

Therefore to test null hypothesis $\alpha=0$ the statistic

$$U=\frac{N-p-q-1}{q}(R-1)$$

can be used as a variance ratio with q and $(N-p-q-1)$ degrees of freedom as observed in earlier papers. The unconditional distribution of w is not so as shown below.

The joint distribution (3.3) of S and R, when $\beta=0$, is

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

*In an earlier paper I have suggested the use of the difference $(n_1+n_2-2)(M_{p+q}-M_p)$ as χ^2 on q degrees of freedom to test the significance of the additional q characters even when $\beta\neq 0$ provided the dispersion matrix is estimated on a large number of degrees of freedom (Rao, 1948a, p.66.)

Making the transformation

$$\frac{1-R}{RS} = w, \qquad S = S$$

so that $dR\,dS = R^2 S\,dw\,dS$ the joint distribution of w and S becomes

$$(\beta_1 \beta_2)^{-1} S^{\frac{N-p+q-1}{2}-1} (1-S)^{\frac{p}{2}-1} e^{-c\alpha^2 S/2}$$

$$\times \sum_{t=0}^{\infty} \frac{m}{t!} \left(\frac{c\alpha^2 S^2}{2}\right)^t w^{\frac{q}{2}+t-1} (1+wS)^{-\frac{N-p-1}{2}-t} dw\,dS$$

where $\beta_1 = \beta\left(\frac{N-p-1}{2}, \frac{p}{2}\right)$, $\beta_2 = \beta\left(\frac{N-p-q-1}{2}, \frac{q}{2}\right)$

and $m = \Gamma\left(\frac{N-p-1}{2}+t\right)\Gamma\left(\frac{q}{2}\right) / \Gamma\left(\frac{N-p-1}{2}\right)\Gamma\left(\frac{q}{2}+t\right)$

The statistic w may be replaced by W connected by the relation

$$w = \frac{W}{1-W}$$

so that W varies from 0 to 1. The joint distribution W and S is given by

$$(\beta_1 \beta_2)^{-1} S^{\frac{N-p+q-1}{2}-1} (1-S)^{\frac{p}{2}-1} e^{-c\alpha^2 S/2}$$

$$\times \sum_{t=0}^{\infty} \frac{m}{t!} \left(\frac{c\alpha^2 S^2}{2}\right)^t W^{\frac{q}{2}+t-1} (1-W)^{\frac{N-p-q-1}{2}-1}$$

$$\times (1-W\overline{1-S})^{-\frac{N-p-1}{2}-t} dW\,dS \qquad (3.7)$$

The distribution of W is obtained by integrating the above expression with respect to S from 0 to 1. I prefer to retain the distribution of W in the integral form instead of expanding the functions $e^{-c\alpha^2 S/2}$ and $(1-W\overline{1-S})^{-\frac{N-p-1}{2}-t}$ in power series of S and $(1-S)$ and integrating out term by term. Some simpler forms obtained as approximations are under consideration and they will be published in another paper.

What is of interest is the null distribution of the statistic W. Putting $\alpha=0$, the joint distribution of W and S becomes

$$(\beta_1\beta_2)^{-1}\, S^{\frac{N-p+q-1}{2}-1}\,(1-S)^{\frac{p}{2}-1}\,(1-W)^{\frac{N-p-q-1}{2}-1}$$

$$\times\; W^{\frac{q}{2}-1}\,(1-W\sqrt{1-S})^{-\frac{N-p-1}{2}}\,dW\,dS$$

Using the following expansion

$$(1-W\sqrt{1-S})^{-\frac{N-p-1}{2}}=\sum_{r=0}^{\infty}\frac{\Gamma\left(\frac{N-p-1}{2}+r\right)}{\Gamma\left(\frac{N-p-1}{2}\right)}\,\frac{W^r(1-S)^r}{r!}$$

and integrating out term by term the distribution of W becomes

$$\sum_{r=0}^{\infty}\frac{\beta\left(\frac{N-p+q-1}{2},\frac{p}{2}+r\right)}{\beta_1\beta_2}\,\frac{\Gamma\left(\frac{N-p-1}{2}+r\right)}{\Gamma\left(\frac{N-p-1}{2}\right)}\,\frac{1}{r!}$$

$$\times\; W^{\frac{q}{2}+r-1}\,(1-W)^{\frac{N-p-q-1}{2}-1}\,dW$$

which can be thrown into the neat form.

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

$$\times\; {}_2F_1\left(\frac{p}{2},\frac{N-p-1}{2};\frac{N+q-1}{2},W\right)\,dW \qquad (3.8)$$

The percentage points of this distribution are under preparation in the statistical Laboratory at Calcutta. The non-null distributions of W and R are very complicated so that it will not be easy to compare the power functions connected with these two test criteria. In a later subsection it was shown that of the estimates of α based on R and the weighted difference of M_{p+q} and M_p, the latter has a smaller variance. So far as terms of the order N^{-2} are concerned the estimates based on $w=M_{p+q}-M_p$ and the weighted difference have the same variance. This indicates the possibility of the test based on w or W being more powerful, than that based on the ratio R. Thus when $\beta=0$, it is possible to increase the efficiency of the test by using the alternative statistic W.

It may be mentioned that this test is available when and only when β is known to be zero *a priori* and the measurements on the first p characters are random variables.

The effect of increasing the number of characters : Using the non-null distribution of the D^2-statistic one can calculate the power of the test for any given sizes of the samples (n_1, n_2) and distance Δ. For the same sample sizes let the power functions corresponding to Δ_p and Δ_{p+q} the distances based on p and $p+q$ characters be denoted by $P(\Delta_p)$ and $P(\Delta_{p+q})$. The q more additional characters increase the efficiency of the test only if

$$P(\Delta_p) < P(\Delta_{p+q}) \tag{3.9}$$

Extensive tables of the power function of the D^2-statistic are necessary to examine the above problem in detail. Some tentative conclusions can be arrived at with the help of the limited tables prepared by Tang(1938). The following Table 2 corrtructed from Tang's tables gives the power of the D^2 test for different values of $N=n_1+n_2$, p, the number of charaters and $\phi^2=c\Delta_p^2/(p+1)$, $c=n_1 n_2/n_1+n_2$.

TABLE 2

POWER FUNCTION FOR THE D^2-STATISTIC

p \ N	$\phi=1$				$\phi=1.5$				$\phi=2$			
	16	20	24	28	16	20	24	28	16	20	24	28
1	.26	.27	.27	.28	.51	.52	.53	.53	.75	.76	.77	.78
2	.28	.28	.28	.29	.53	.56	.57	.58	.80	.82	.83	.84
3	.27	.29	.30	.31	.56	.60	.62	.63	.82	.86	.87	.89
4	.27	.30	.32	.33	.57	.62	.65	.67	.84	.88	.90	.91
5	.27	.30	.33	.34	.57	.64	.68	.71	.84	.90	.92	.94
6	.26	.31	.34	.36	.56	.65	.70	.73	.83	.90	.94	.9
7	.25	.30	.34	.37	.54	.65	.71	.75	.81	.91	.94	.96
8	.24	.30	.34	.37	.50	.64	.72	.76	.78	.90	.95	.96

Since the power of the test increases with increase in ϕ it immediately follows that for any given set of characters to be used for discrimination and the total sample size N, the power is a maximum for the largest value of c. This means it is profitable to divide the samples *equally* between the two populations.

From the above table it is seen that, for given sample sizes, the power increases with increase in the number of characters up to a certain stage (underlined in the above table) and then decreases provided ϕ remains constant *i.e.*,

$$\frac{\Delta_1^2}{2} = \frac{\Delta_2^2}{3} = \ldots \tag{3.10}$$

With increase in the size of the samples the decrease in the loss of power with the increase in the number of characters, in the above situation 3.10, occurs only after a

sufficiently large number of characters is included. One can safely conclude that so long as the newly added character, $(p+1)$th is such that

$$\frac{\Delta^2_{p+1}}{p+2} \geq \frac{\Delta^2_p}{p+1}$$

the power increases provided the total sample size (N) is not very small and p is not very large. Even when N is as small as 24 or 28 it is seen from Table 2 that the power increases even beyond the value $p=8$ for the values of ϕ considered above.

In the example considered in Section 2, $D^2_1=.4614$ (due to Femur) and $D^2_2=.4777$ so that the increase in D^2 by the inclusion of Humerus is .0163. Such a small increase is not of value in samples of sizes 20 and 27 for the two populations. In fact the power of the test including Humerus gets reduced. With ten more observations on the total and equal distribution of samples the inclusion of Humerus would have been useful.

Gain in efficiency due to covariance : Cochran and Bliss(1948) considered the problem of testing the significance of the population distance when it is known that a subset p of characters by themselves do not afford any discrimination. The test is the same as that used for judging the additional contribution to distance between two populations due to a set of q characters when the differences due to the first p characters are eliminated. An alternative to this test is to ignore the subset of characters which afford no discrimination and use the D^2-statistic based on the q other characters alone. If γ^2 is the population parameter giving the square of the distance with respect to q characters then the distribution of $S'=1/M_q$ is

$$e^{-c\gamma^2/2} \sum_{t=0}^{\infty} \frac{1}{t!} \left(\frac{c\gamma^2}{2}\right)^t B\left(\frac{N-q-1}{2}, t+\frac{q}{2}\right) dS' \qquad (3.11)$$

The distribution of R which takes into account the effect of the first p characters is obtainable from the distribution (3.4) by putting $\beta=0$. This is

$$\sum_{t=0}^{\infty} \left\{ \int_0^1 B\left(\frac{N-p-1}{2}, \frac{p}{2}\right) e^{-c\alpha^2 S/2} \left(\frac{c\alpha^2 S}{2}\right)^t dS \right\} B\left(\frac{N-p-q-1}{2}, t+\frac{q}{2}\right) dR$$

The conditional power of the test associated with R is

$$\int_0^{R_0} e^{-c\alpha^2 S/2} \sum_{t=0}^{\infty} \frac{1}{t!} \left(\frac{c\alpha^2 S}{2}\right)^t B\left(\frac{N-p-q-1}{2}, t+\frac{q}{2}\right) dR$$

where R_0 is the 5% significant value of R for testing the null hypothesis $\alpha=0$.

If this is represented by

$$P\left(\frac{N-p-q-1}{2}, q, \alpha^2 S\right)$$

then the average power or the unconditional power of the R test is

$$\int_0^1 P\left(\frac{N-p-q-1}{2}, q, \alpha^2 S\right) B\left(\frac{N-p-1}{2}, \frac{p}{2}\right) dS$$

$$= P(q \mid p, \alpha^2) \text{ say.}$$

The power associated with the test S′ of (3.11) is

$$P\left(\frac{N-q-1}{2}, q, \gamma^2\right)$$

In some situations what is of importance is the conditional power so that the R test will be more efficient if

$$P\left(\frac{N-p-q-1}{2}, q, \alpha^2 S\right) > P\left(\frac{N-q-1}{2}, q, \gamma^2\right)$$

If p is small compared to N the power functions differ only in values of $\alpha^2 S$ and γ^2. In this case the above inequality can be replaced by $\alpha^2 S > \gamma^2$. α^2 is always not less than γ^2, the distance based on $p+q$ characters being not less than that, on q characters. The maximum value of S is unity so that $\alpha^2 S$ could be less than γ^2 if α^2 is not sufficiently greater than γ^2. As observed earlier, the most favourable situation is when a high value of S is observed or the two samples are deliberately chosen to provide a high value of S which means a smaller distance based on the p chatacters.

To judge the general efficiency of the R test we can only compare the average power $P(q \mid p, \alpha^2)$ with $P\left(\frac{N-q-1}{2}, q, \gamma^2\right)$ the power of the S′ test. The former will be more efficient if

$$P(q \mid p, \alpha^2) > P\left(\frac{N-q-1}{2}, q, \gamma^2\right)$$

A numerical study of the above relationship can be made when once extensive tables of the power function of D^2 are prepared.

Bias in the D^2-statistic: The quantity S defined above is connected with D_p^2 by the relation

$$S = \frac{1}{M_p} = \frac{1}{1 + \frac{c}{N-2} D_p^2}$$

Knowing the moments of M_p or $1/S$, the moments of D_p^2 can be easily derived. Using the distribution of S given in (3.2) we find

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

$$= e^{-c\beta^2/2} \sum_{r=0}^{\infty} \frac{1}{r!} \left(\frac{c\beta^2}{2}\right)^r \frac{\Gamma\left(\frac{N-p-1}{2}-t\right)\Gamma\left(\frac{N-1}{2}+r\right)}{\Gamma\left(\frac{N-1}{2}-t+r\right)\Gamma\left(\frac{N-p-1}{2}\right)}$$

$$= e^{-c\beta^2/2} \frac{\Gamma\left(\frac{N-p-1}{2}-t\right)\Gamma\left(\frac{N-1}{2}\right)}{\Gamma\left(\frac{N-p-1}{2}\right)\Gamma\left(\frac{N-1}{2}-t\right)} {}_1F_1\left(\frac{N-1}{2},\ \frac{N-1}{2}-t,\ \frac{c\beta^2}{2}\right)$$

Applying Kummer's transformation

$${}_1F_1(\rho, \gamma, x) = e^x\, {}_1F_1(\gamma-\rho, \gamma, -x)$$

the above expression becomes

$$E(M_p^{-t}) = \frac{\Gamma\left(\frac{N-p-1}{2}-t\right)\Gamma\left(\frac{N-1}{2}\right)}{\Gamma\left(\frac{N-p-1}{2}\right)\Gamma\left(\frac{N-1}{2}-t\right)} \quad {}_1F_1\left(-t,\ \frac{N-1}{2}-t,\ -\frac{c\beta^2}{2}\right)$$

Putting $t=1$,

$$E(M_p) = E\left(1+\frac{c}{N-2} D_p^2\right)$$

$$= \frac{N-3}{N-p-3}\left\{1+\frac{c\beta^2}{N-3}\right\}$$

Re-arranging we find

$$E(D_p^2) = \frac{N-2}{N-p-3}\left\{\beta^2+\frac{p}{c}\right\}$$

An unbiased extimate of β^2 is

$$\frac{N-p-3}{N-2} D_p^2 - \frac{p}{c} \tag{3.12}$$

so that D_p^2 overestimates the population parameter β^2. When p is small and N is large the above expression does not differ very much from D_p^2. The bias is a minimum for a given N if samplo sizes are equal in the two populations for in this case $c=n_1n_2/(n_1+n_2)$ is a maximum. No correction is needed if a number of D^2's arising out of a given number of groups whose sample sizes do not vary much have to be compared.

It is of some interest to determine an unbiased estimate of β^2 in the case when the mean values are obtained from samples of sizes n_1 and n_2 and variances and covariances are estimated on a large number of degrees of freedom. If f is the degrees of freedom for the variances and covariances, the exact distribution of

$$S = \frac{1}{N_p} = \frac{1}{1+\frac{c}{f}D_p^2}$$

is obtained by replacing N by $(f+2)$ in the distribution (3.2). Hence an unbiased estimate of β^2 is given by

$$\frac{f-p-1}{f}\,D_p^2 - \frac{p}{c}$$

When f is large the estimate can be taken as simply

$$D_p^2 - \frac{p}{c}$$

The correction need not be used if the c's for the various D^2's to be compared do not vary much.

The estimation of the additional distance: Having estimated β^2 it may be of interest to estimate the additional contribution to the square of the distance when q more characters are added to the first p. From the results derived above an unbiased estimate of $\alpha^2+\beta^2$ is

$$\frac{N-p-q-3}{N-2}\,D^2_{p+q} - \frac{p+q}{c}$$

Subtracting from this the unbiased estimate (3.12) of β^2 we derive the unbiased estimate of α^2

$$\frac{(N-p-q-3)D^2_{p+q}-(N-p-3)D^2_p}{N-2} - \frac{q}{c} \tag{3.13}$$

which is approximately

$$D^2_{p+q}-D^2_p - \frac{q}{c}$$

if N is large. A simple method of computing these quantities is given in Appendix 1.

An interesting case considered by Cochran and Bliss (1948) is the estimate of the distance based on $p+q$ characters when it is known that the first p characters by themselves contribute nothing to the distance *i.e.*, the mean values are the same for the first p characters in the two populations or $\beta=0$. The problem is, then, that of estimating α^2.

One estimate is

$$a^2_1 = \frac{(N-p-q-3)D^2_{p+q}-(N-p-3)D^2_p}{N-2} - \frac{q}{c}$$

as derived in equation (3.13).

Another estimate is derivable from E(1/R).

$$E\left(\frac{1}{R}\right) = E\left(\frac{M_{p+q}}{M_p}\right) = \int\int \frac{1}{R} P(R,S)\, dR\, dS$$

where $P(R, S)\, dRdS$ is the joint probability distribution of R and S derived in (3.3).

$$E\left(\frac{1}{R}\right) = \frac{N-p-3}{N-p-q-3} + \frac{c\alpha^2}{N-p-q-3} E(S)$$

$$= \frac{N-p-3}{N-p-q-3} + \frac{c\alpha^2}{N-p-q-3}\frac{N-p-1}{N-1}\, {}_1F_1\left(1, \frac{N+1}{2}, -\frac{c\beta^2}{2}\right)$$

$$= \frac{N-p-3}{N-p-q-3} + \frac{c\alpha^2}{N-p-q-3}\frac{N-p-1}{N-1}, \text{ if } \beta=0.$$

An unbiased estimate of α^2 when $\beta=0$ is given by

$$a^2_2 = \left\{\frac{1}{R} - \frac{N-p-3}{N-p-q-3}\right\}\frac{(N-p-q-3)(N-1)}{c(N-p-1)} \tag{3.14}$$

To determine which of the two estimates a^2_1, (3.13) and a^2_2, (3.14) is better we compare their variances. The expectation of $1/R^2$ derivable from the joint distribution of S and R, is

$$E\left(\frac{1}{R^2}\right) = \frac{4}{(N-p-q-3)(N-p-q-5)}\left\{\frac{(N-p-3)(N-p-5)}{4} + \frac{N-p-3}{2} c\alpha^2 E(S) + \frac{c^2\alpha^4}{4} E(S^2)\right\}$$

$$= \frac{4}{(N-p-q-3)(N-p-q-5)}\left\{\frac{(N-p-3)(N-p-5)}{4} + \frac{(N-p-3)(N-p-1)}{N-1}\cdot\frac{c\alpha^2}{2} + \frac{(N-p+1)(N-p-1)}{(N+1)(N-1)}\cdot\frac{c^2\alpha^4}{4}\right\}$$

when $\beta=0$. To find the variance of a_1^2 we observe that

$$a_1^2 = \frac{(N-p-q-3)M_{p+q}-(N-p-3)M_p}{c}$$

so that its variance can be derived with the help of the following expressions

$$E(M_p^2) = \frac{(N-3)(N-5)}{(N-p-3)(N-p-5)}$$

$$E(M_{p+q}^2) = \frac{(N-3)(N-5)+2cx^2(N-3)+c^2x^4}{(N-p-q-3)(N-p-q-5)}$$

$$E(M_{p+q}M_q) = E(1/RS^2)$$

$$= \frac{N-p-3}{N-p-q-3}\ E(M_p^2) + \frac{cx^2(N-3)}{(N-p-q-3)(N-p-3)}$$

Retaining up to the terms of order $1/N^2$ the expressions for variances are

$$V(a_2^2) = \frac{2x^4+4xx^2}{N} + \frac{2(2p+q+5)x^4+4x(p+q+2)x^2+2qx^2}{N^2}$$

$$V(a_1^2) = \frac{2x^4+4xx^2}{N} + \frac{2(p+q+5)x^4+4x(p+q+2)x^2+2qx^2}{N^2}$$

where $x=N/c$. The variances do not differ in terms of the order $1/N$, so that in large samples both the statistics are equally efficient, the common asymptotic variance being

$$\frac{2x^4+4xx^2}{N}$$

The actual difference between the variances retaining only terms up to the order $1/N^2$ is

$$\frac{2px^4}{N^2}$$

so that in *moderately large samples* the statistic a_1^2 being less variable, should be preferred to a_2^2.

4. Some Examples of dimensional convergence

Consider the correlation matrix

$$\begin{pmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ . & . & . & \dots & . \\ \rho & \rho & \rho & \dots & 1 \end{pmatrix}$$

of q characters with differences $d_1, \dots, d_q$ in the mean values of two populations expressed in their respective standard deviation units. In this case

$$D^2_p = \frac{1}{1-\rho}\{\Sigma d^2_1 - \rho(\Sigma d_1)^2/(1+\overline{q-1}\,\rho)\}$$

$$= q\left\{\frac{\sigma^2_d}{1-\rho} + \frac{\bar{d}^2}{1+(q-1)\rho}\right\}$$

where $\bar{d}$ is the average value of $d_1, \dots d_q$ and σ^2_d, their variance.

As $q \to \infty$ the above expression tends to

$$\frac{\text{Limit } q\sigma^2_d}{1-\rho} + \frac{\bar{d}^2}{\rho}$$

Unless the characters are such that the variance of the differences* in mean values falls rapidly with increase in the number of characters, the above expression does not tend to a finite limit. If all the differences are of the same magnitude then the above expression becomes $\bar{d}^2/\rho$ which is the maximum value of D^2 attainable.

Consider the correlation matrix

$$\left(\begin{array}{c:c} A & B \\ \hdashline B & C \end{array}\right)$$

where

$$A = \begin{pmatrix} 1 & \rho_1 & \rho_1 & \dots & \rho_1 \\ \rho_1 & 1 & \rho_1 & \dots & \rho_1 \\ . & . & . & \dots & . \\ \rho_1 & \rho_1 & \rho_1 & \dots & 1 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & \rho_2 & \rho_2 & \dots & \rho_2 \\ \rho_2 & 1 & \rho_2 & \dots & \rho_2 \\ . & . & . & \dots & . \\ \rho_2 & \rho_2 & \rho_2 & \dots & 1 \end{pmatrix}$$

and

$$B = \begin{pmatrix} \rho_3 & \rho_3 \dots \rho_3 \\ \rho_3 & \rho_3 & \rho_3 \\ . & . & . \\ \rho_3 & \rho_3 \dots \rho_3 \end{pmatrix}$$

* Only the standardised differences are being considered.

A is the corelation matrix of q characters with $d_1, \ldots, d_q$ as the standardised differences in the population means, C is that of p characters with differences $\delta_1, \ldots, \delta_p$ and B is the matrix of cross correlations.

In this case D^2 is given by

$$\alpha\Sigma d^2_1 + \alpha'\Sigma\delta^2_1 + \beta(\Sigma d)^2 + \beta'(\Sigma\delta)^2 + 2\omega(\Sigma d)(\Sigma\delta)$$

where

$$\alpha = \frac{1}{1-\rho_1}, \quad \alpha' = \frac{1}{1-\rho_2}$$

$$\beta = \frac{p\rho^2_3 - \rho_1(1+\overline{p-1}\,\rho_2)}{\gamma(1-\rho_1)}, \quad \beta' = \frac{q\rho^2_3 - \rho_2(1+\overline{q-1}\,\rho_1)}{\gamma(1-\rho_2)}$$

$$\omega = -\frac{\rho_3}{\gamma}, \quad \gamma = (1+\overline{p-1}\rho_2)(1-\overline{q-1}\rho_1) - pq\rho^2_3$$

If d_i are all equal to d and δ_i to δ then D^2 becomes

$$\frac{qd^2(1+\overline{p-1}\rho_2) + p\delta^2(1+\overline{q-1}\rho_1) - 2pqd\delta\rho_3}{(1+\overline{q-1}\rho_1)(1+\overline{p-1}\rho_2) - pq\rho^2_3}$$

If $\rho_1\rho_2 - \rho^2_3 \neq 0$, the above expression as p and $q \to \infty$ tends to

$$\frac{\rho_2 d^2 + \rho_1\delta^2 - 2\rho_3 d\delta}{\rho_1\rho_2 - \rho^2_3}$$

This being the upper limit, D^2 increases very slowly after a certain stage.

5. Summary

(1) Whatever may be the number of characters chosen to discriminate between two populations, it is profitable to divide the samples equally between the two populations.

(2) If the square of the distance increases proportionately with the number of characters, then, except in very small samples, no loss is incurred by the inclusion of extra characters.

(3) To judge the significance of the increase is distance due to the addition of q characters to a set of p chosen characters, it is profitable to choose individuals from the two groups such that on the average, the two series of samples agree as closely as possible in all the p initial characters. Two tests are discussed in this connexion.

(4) Some models have been considered where the distance function tends to a finite limit with increase in the number of characters.

(5) Some computational aspects of the D^2-statistic useful in the above study are illustrated in Appendix 1.

APPENDIX 1. THE COMPUTATIONAL ASPECTS OF THE D^2-STATISTIC

The D^2-statistic is defined by

$$D^2 = \sum\sum s^{ij} d_i d_j$$

where (s^{ij}) is the matrix reciprocal to the dispersion matrix estimated on a certain number of degress of freedom from the data and d_i is the difference between the mean values of the ith character for the two populations.

One way of calculating this is first to transform the original variates into a set of uncorrelated variates and construct the D^2 with respect to the new variables. In the latter case D^2 reduces to a simple sum of squares and because D^2 is invariant under linear transformations of the variates the same value is obtained. A simple method of constructing a linear transformation to make the new variables uncorrelated is given in Appendix 5 of the paper by Mahalanobis, Majumdar and Rao (1949). This method is recommended when a large number of D^2's are to be calculated for a given group of populations having the same dispersion matrix.

A second method is to solve the system of linear equations.

$$\begin{aligned} \lambda_1 s_{11} + \ldots + \lambda_p s_{1p} &= d_1 \\ \cdots \quad \cdots \quad \cdots \\ \lambda_1 s_{p1} + \ldots + \lambda_p s_{pp} &= d_p \end{aligned}$$

and compute the expression

$$D^2 = \lambda_1 d_1 + \ldots + \lambda_p d_p$$

This method is recommended when, in addition to the D^2 value, the compounding coefficients $(\lambda_1, \ldots, \lambda_p)$ of the discriminant function are also needed.

A third way is to calculate the ratio of two determinants and obtain D^2 from the relation

$$1 + D^2 = \frac{|s_{ij} + d_i d_j|}{|s_{ij}|}$$

An alternative method which is more fruitful in studying the various components of D^2 is given by the relation

$$D^2 = -\begin{vmatrix} s_{ij} & \vdots & d' \\ \cdots & \cdots & \cdots \\ d & \vdots & 0 \end{vmatrix} \div |s_{ij}|$$

In this case the matrix in the numerator is reduced by the method of sweep out or pivotal condensation starting from the first pivotal element. The value of the last element of the leading diagonal at any stage of the above process gives the value of D^2 (with a negative sign) based on a number of characters equal to the number of rows or columns swept out. An example is given below.

The dispersion matrix of four variables and the differences in mean values are given below.

dispersion matrix				difference in means
.1953	.0996	.0922	.0331	.030
.0996	.1255	.0472	.0396	2.798
.0922	.0472	.1211	.0252	−.658
.0331	.0396	.0252	.0251	1.080
.030	2.798	−.658	1.080	0

Dividing by the first element .1953 the first row becomes

1	.509985	.472094	.169483	4.761904

Using this as the pivotal row the first element in each of the successive rows is eliminated. Thus multiplying this row by .0996 and subtracting from the second row the first element in the second row is made zero. We obtain the reduced matrix (omitting the elements below the diagonal which are same as those above it).

.074705	.000179	.022719	2.323714
	.077573	.009574	−1.097047
		.010490	.022381
			−4.428571 = $-D_1^2$

where D_1^2 is D^2 based on the first i characters. The next pivotal row and the rows from which the first element is eliminated are

1	.002396	.304116	31.105189
	.077572	.009519	−1.102615
		.012581	.215702
		−4.428571	−72.279565 = $-D_2^2$

The additional contribution to D^2 by including the second character is 72.279565. Reducing further we obtain

	1	.122712	−14.336786
		.011413	.352174
	−4.428571	−72.279565	−15.807960 = $-D_3^2$
		1	30.857260
−4.428571	−72.279565	−15.807960	−10.807124 = $-D_4^2$

$D^2_4 = 103.383220.$

To test the significance of D_4^2 we use

$$F = \frac{n_1 n_2 (n_1+n_2-1-4)}{(n_1+n_2)(n_1+n_2-2)} \frac{D_4^2}{4}$$

as a variance ratio with 4 and (n_1+n_2-5) degrees of freedom (see Rao, 1948a). In the above formula n_1 and n_2 are sample sizes. If it is desired to test whether the last two characters supply additional information for discrimination, the statistic is

$$\frac{n_1+n_2-5}{2} \frac{n_1 n_2 (D_4^2 - D_3^2)}{(n_1+n_2)(n_1+n_2-2) - n_1 n_2 D_3^2}$$

which can be used as a variance ratio with 2 and (n_1+n_2-5) degrees of freedom. The quantities D_3^2 and $D_4^2 - D_3^2$ are easily made available in the above scheme of computation.

The above method is useful when only a few D^2's have to be computed and the significance of the additional contribution due to a group of characters has to be tested.

Appendix 2. The calculation of Wilk's Λ criterion

In the case of discrimination of many groups with multiple variates an overall test is provided by Wilk's criterion defined by

$$\Lambda = \frac{|\ W\ |}{|\ T\ |}$$

where W is the matrix of sum of squares and products within groups and T is the matrix of total sum of squares and products. To evaluate the determinants it is convenient to follow the method of pivotal condensation. If $w_1, w_2, \ldots, w_p$ are the pivotal elements* for the matrix W then

$$|\ W\ | = w_1 w_2 \ldots w_p$$

If $t_1, \ldots, t_p$ are the pivotal elements for the matrix T then

$$|\ T\ | = t_1 t_2 \ldots\ t_p$$

The Wilk's criterion constructed out of the first i characters is

$$\Lambda = \frac{w_1 w_2 \ldots w_i}{t_1 t_2 \ldots t_i}$$

To test for further discrimination supplied by the remaining $(p-i)$ characters the

* The first element which is made unity in any pivotal row is called a pivotal element.

statistic to be used is

$$\Lambda'_p = \frac{w_{i+1}\, w_{i+2} \ldots w_p}{l_{i+1}\, l_{i+2} \ldots l_p}$$

The statistic for all the characters is

$$\Lambda_1 \Lambda'_p = \Lambda_p = \frac{w_1 w_2 \ldots w_p}{l_1 l_2 \ldots l_p}$$

In a previous paper by the author (Rao, 1948a) a different method was suggested to calculate Λ'_p. The first step is to obtain the within and total matrices of squares and products for $p-i$ characters after eliminating the effect of first i characters. If these are denoted by $W(p-i,i)$ and $T(p-i,i)$ then

$$\Lambda'_p = \frac{|W(p-i,i)|}{|T(p-i,i)|}$$

This elaborate procedure is not needed when only Λ'_p has to be calculated. The method of reducing the determinants by pivotal elements is easier. If the object is to construct the canonical variances corresponding to the residual matrices then the actual matrices $W(p-i,i)$ and $T(p-i,i)$ will be needed. For tests with the Λ criterion reference may be made to Bartlett (1947) and Rao (1948a).

Appendix 3. A property of the β—distribution

If x and y are two independent variables following the β—distributions

$$x^{a-1}(1-x)^{b-1}/\beta(a,b)$$

$$y^{c-1}(1-x)^{d-1}/\beta(c,d)$$

then the product $z=xy$ is also distributed in the β form $B(c,b+d)$ provided $a=c+d$.

Proof : The joint distribution of x and y is

$$\text{Const.}\quad x^{a-1}(1-x)^{b-1}y^{c-1}(1-y)^{d-1}dxdy$$

Put $x=x, z=xy$ so that $\partial(x,y)/\partial(x,x)=1/x$. The joint distribution of x and z is

$$\text{Const.}\quad x^{a-c-d}(1-x)^{b-1}z^{c-1}(x-z)^{d-1}dxdz$$

$$=\text{Const.}\quad (1-x)^{b-1}z^{c-1}(x-z)^{d-1}dxdz, \text{ since } a=c+d$$

Integrating this over x from z to 1, the probability density of z is obtained as

$$z^{c-1}(1-z)^{b+d-1}/\beta(c,b+d)$$

If $x_1, x_2, \ldots$ are independent variates with distributions

$$B(a_1, b_1)dx_1, \quad B(a_2, b_2)dx_2 \ldots\ldots\ldots$$

then the distribution of the product $x_1 x_2 \ldots = x$ is

$$B(a_1, b_1+b_2+b_3\ldots)\, dx$$

provided $a_2 = a_1 + b_1$, $a_3 = a_2 + b_2 \ldots\ldots$ This follows from the above result.

BIBLIOGRAPHY

BARTLETT, M. S. (1947). Multivariate analysis. *J. Roy. Statist. Soc. Suppl.* 9, 76.

——— (1948). Internal and External factor analysis. *British Jour. of Psychology.* (Statistical section). 1, 73.

BOSE, R. C. & ROY, S. N. (1938). The distribution of the Studentised D²-statistic. *Sankhya*, 4, 19.

COCHRAN, W. C. and BLISS, C. I. (1948). Discriminant functions with covariance. *Ann. Math. Statist.* 09, 151

FISHER, R. A. (1936). The use of multiple measurements in taxonomic problems. *Ann. Eugen. Lond.*, 7, 179.

——— (1938). The statistical utilization of multiple measurements. *ibid*, 8, 376.

GEARY, R. C. (1948). Studies in relations between economic time series. *J. Roy Statist. Soc.* series B, 10,140.

HSU, P. L. (1938). Notes on Hotelling's generalized T. *Ann. Math. Statist.*, 9, 231.

MAHALANOBIS, P. C. (1936). On the generalized distance in statistics. *Proc. Nat. Inst. Sci.* (India), 12, 49.

MAHALANOBIS, P. C., BOSE, R. C. and ROY, S. N. (1937). Normalisation of statistical variates and the use of rectangular coordinates in the theory of sampling distributions. *Sankhya*, 3, 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. (1946a). On the linear combination of observations and the general theory of least squares. *Sankhya*, 7, 237.

——— (1946b). Tests with discriminant functions in multivariate analysis. *Sankhya*, 7, 407.

——— (1948a). Tests of significance in multivariate analysis. *Biometrika*, 35, 58.

——— (1948b). The utilization of multiple measurements in problems of biological classification. *J. Roy. Statist. Soc.* Series B, 10, 159.

RAO, C. R. and SLATER, P. (1949). Statistical treatment of neurotic cases. *British Jour. of Psychology* (Statistical Section), in press.

TANG, P. C. (1938). The power function of the analysis of variance tests with tables and illustrations of their use. *Statist. Res. Mem.* 2, 126.

TINTNER, G. (1946). Some applications of multivariate analysis to economic data. *Econometrica*, 14, 5.