155

THE STATISTICAL UTILIZATION OF MULTIPLE MEASUREMENTS

Author's Note (CMS 33.375a)

Papers 155 and 175 attempt to bring under a common point of view diverse researches, of which the most important had been initiated by Hotelling in the United States and by Mahalanobis in India. The author's own researches had approached essentially the same problems by a technique known as discriminant functions. The results have here been compared in a common notation, and the first steps taken to advance the theory of discriminant functions so far as to test their significance and the collinearity or coplanarity of observed aggregates.

The slip alluded to in Paper 175 has been corrected in the present edition of Paper 155, and the treatment brought into line with that of Paper 175.

 Annals of Eugenics, 8: 376-386, (1938).

THE STATISTICAL UTILIZATION OF MULTIPLE MEASUREMENTS

By R. A. FISHER

I. Introductory

It has been shown (Barnard, 1935; Fairfield Smith, 1936; Fisher, 1936) that a set of multiple measurements may be used to provide a discriminant function, linear in the observations, having the property that, better than any other linear function, it will discriminate between any chosen classes such as taxonomic species, the two sexes, plants giving more or less desirable progeny, and so on. Its use in metrical psychology has been illustrated by Wallace & Travers (1938).

In discussing the application of this process to a taxonomic problem, I was led to point out its formal analogy with the process of fitting an equation of multiple regression. The type of problem involved is also closely related to problems earlier discussed, on the one hand, by P. C. Mahalanobis (1927, 1930, 1936) and on the other by H. Hotelling (1931). It may, therefore, be of some value to show the connexion between these three different lines of work, and to distinguish between the objects for which they were developed.

If we have samples of N_1 and N_2 objects respectively, and make p measurements $x_1, \ldots x_p$ on each, the analogy between the calculations of a discriminant function (written now with upper affices)

$$X = b^1 x_1 + b^2 x_2 + \ldots + b^p x_p,$$

which shall best distinguish objects of one class from those of another, and the procedure of multiple regression, is brought out by introducing a formal dependent variate y, which is given the value $N_2/(N_1+N_2)$ for objects of the first, and $-N_1/(N_1+N_2)$ for objects of the second class.

These conventional values ensure that the average values of y in the two classes shall differ by unity, and that

$$S(y) = 0,$$

$$S(y^2) = \frac{N_1 N_2}{N_1+N_2} = \lambda^2,$$

where the summation is taken over all the objects observed.

The multiple regression equation for predicting the value of y from observed values $x_1, \ldots, x_p$ is now of the form

$$Y = \sum_{i=1}^{p} b^i (x_i - \bar{x}_i),$$

where the regression coefficients, $b^1, \ldots, b^p$, are given by the equations

$$\sum_{j=1}^{p} s'_{ij} b^j = \lambda^2 d_i,$$

where $$s'_{ij} = S(x_i - \bar{x}_i)(x_j - \bar{x}_j),$$
x_i stands for any of the variates $x_1, \ldots, x_p$ and d_i for the difference between the mean of x_i in the first sample and that in the second.

To demonstrate that we may take the coefficients b obtained from these regression equations as the coefficients of the discriminant function, note that
$$s'_{ij} = ns_{ij} + \lambda^2 d_i d_j,$$
where s_{ij} stands for the mean product of the variates x_i and x_j taken within the two samples, and n for the degrees of freedom within samples.

Substituting this expression for s'_{ij} in the regression equations, they take the form
$$\sum_{j=1}^{p} (ns_{ij} + \lambda^2 d_i d_j) b^j = \lambda^2 d_i,$$
whence $$n \sum_{j=1}^{p} s_{ij} b^j = \lambda^2 d_i \left(1 - \sum_{j=1}^{p} b^j d_j\right),$$
showing that the coefficients obtained differ only by the constant factor
$$\lambda^2\{1 - \Sigma(bd)\}$$
from the solution of the equations $$n \sum_{j=1}^{p} s_{ij} b^j = d_i$$
obtained (Fisher, 1936, p. 181) for the coefficients of the discriminant function.

II. The analysis of variance

By fitting the regression equation the variation observed in the variate y has been analysed in two portions. The sum of the products of the regression coefficients, b, and the right-hand side of the regression equation, $\lambda^2 d$, is
$$\lambda^2 \Sigma(bd),$$
and this is the portion accounted for by regression, out of the total λ^2. Consequently we have the analysis

Degrees of freedom	Sum of squares
p	$\lambda^2\Sigma(bd)$
N_1+N_2-p-1	$\lambda^2\{1-\Sigma(bd)\}$
N_1+N_2-1	λ^2

If R stand for the multiple correlation of y with $x_1, \ldots, x_p$, evidently
$$R^2 = \Sigma(bd).$$
This same quantity is the difference between the mean values of X in the two samples observed.

The table of the analysis of variance suggests, though by itself it does not demonstrate, that the significance of R^2 could be tested by applying the ordinary z test to the analysis. Ordinarily, in multiple regression the population postulated has a normal distribution for y for each set of values $x_1 \ldots x_p$. The distribution of the independent variates is then irrelevant. The population postulated in our present problem has fixed values of y, but a simultaneous normal distribution for $x_1, \ldots, x_p$. Hotelling's earlier work shows, however, that the test of significance is exactly that which the analysis of variance suggests.

III. Hotelling's test of significance

The title of Hotelling's paper (1931) shows that he was not concerned with estimates, but with a test of significance. The "Generalization of 'Student's' ratio" at which he arrived is derived from the matrix s_{ij} of dispersion within samples. In connexion with this he uses the vector of differences with a factor reducing it to the precision of a single observation; in our notation

$$\xi = \lambda d.$$

If s^{ij} stand for the element corresponding with s_{ij} in the reciprocal matrix, then Hotelling chooses a form invariant for all linear transformations of $x_1, \ldots, x_p$, and puts

$$T^2 = \sum_{i=1}^{p}\sum_{j=1}^{p} s^{ij}\xi_i\xi_j,$$

where $n\ (=N_1+N_2-2)$ is the number of degrees of freedom within samples.

Now, from the equation

$$n\sum_{j=1}^{p} s_{ij}b^j = \lambda^2(1-R^2)d_i$$

it follows that

$$nb^i = \lambda^2(1-R^2)\sum_{j=1}^{p} s^{ij}d_j$$

$$= \lambda(1-R^2)\sum_{j=1}^{p} s^{ij}\xi_j.$$

Multiplying by d_i and adding, these equations give

$$\Sigma bd = (1-R^2)\,T^2/n,$$

or

$$T^2 = nR^2/(1-R^2).$$

If we calculate the z test of significance from the analysis of variance, we find

$$z = \tfrac{1}{2}\log\frac{R^2}{p} - \tfrac{1}{2}\log\frac{1-R^3}{N_1+N_2-p-1}.$$

Substituting for R in terms of T, and for N_1+N_2 in terms of n, this is

$$z = \tfrac{1}{2}\log\frac{T^2}{pn\left(1+\dfrac{T^2}{n}\right)} + \tfrac{1}{2}\log\left(1+\frac{T^2}{n}\right)(n-p-1)$$

$$= \tfrac{1}{2}\log T^2(n-p+1) - \tfrac{1}{2}\log np,$$

with degrees of freedom $n_1 = p$, $n_2 = n-p+1$, in accordance with the test of significance given by Hotelling (1931, p. 377).

IV. Mahalanobis' generalized distance

The test appropriate for the significance of the discriminant function, that is for significant contradiction of the hypothesis that the samples are from populations undifferentiated in respect of the variates x_1 to x_p, was thus given by Hotelling so early as 1931. Naturally, the scalars T and R give no indication of the direction in p-space in which the two samples are most distinct; they do, however, indirectly measure the distance, or the extent to which the two sets of multiple measurements differ. This is the object of a third series of researches initiated by Mahalanobis in 1927.

If σ_{ij} is a typical element of the dispersion matrix of the populations sampled, and σ^{ij} is the corresponding element of the reciprocal matrix, the property of the population of which Mahalanobis proposes an estimate is

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

where δ stands for the difference in each variate between the population means.

This resembles Hotelling's test of significance in being invariant for all linear transformations of the variates x; evidently also it is only zero if δ_i vanishes for all values of i. It differs from Hotelling's form in being a population parameter capable of estimation. The factor $1/p$ is a convention due to the fact that Mahalanobis, like Hotelling, was led to investigate the subject through recognizing the shortcomings of the various forms of coefficients of racial likeness which had been used by Pearson and his followers.

The practical estimation of Δ^2 takes two forms appropriate to the cases in which the dispersion matrix is taken as known (as are the variances in one form of Pearson's Coefficient of Racial Likeness), and in which it is estimated from the two samples. The first or "unstudentized" form was investigated by R. C. Bose (1936). He used

$$D^2 = \frac{1}{p} \sum_{i=1}^{p} \sum_{j=1}^{p} \sigma^{ij} d_i d_j - \frac{1}{N_1} - \frac{1}{N_2}.$$

The sampling distribution found by Bose is equivalent to the limiting distribution for the multiple correlation coefficient to which I have called attention (Fisher, 1928). If we consider the distribution of a variate B dependent on a population value, β, in such a way that the frequency element of the distribution is

$$\frac{(\frac{1}{2}B^2)^{\frac{1}{2}(p-2)}}{\{\frac{1}{2}(p-2)\}!} e^{-\frac{1}{2}B^2-\frac{1}{2}\beta^2} \left\{1 + \frac{1}{p}\frac{\beta^2 B^2}{2} + \frac{1}{p(p+2)}\frac{\beta^2 B^2}{2.4.} + \dots\right\} d(\tfrac{1}{2}B)^2, \qquad \text{(B)}$$

then the distribution of the multiple correlation coefficient, R, calculated from a large sample from a population having true correlation ρ, will be found by substituting

$$\beta^2 = n\rho^2, \quad B^2 = nR^2,$$

the distribution being exact when n is increased indefinitely, and p is the number of independent variates.

The distribution of the unstudentized statistic D^2 is found equally by making the substitutions

$$\beta^2 = \lambda^2 p \Delta^2,$$
$$B^2 = \lambda^2 p D^2 + p,$$

where λ^2 stands for $N_1 N_2/(N_1 + N_2)$.

The B distribution is also, as was shown in 1928, closely linked with a double Poisson series. A table of the 5 % points is here reproduced from my 1928 paper.

For the case, of greater practical importance, in which the dispersion matrix within samples is not known in advance, but is replaced by s_{ij} obtained by pooling the sums of squares and products from the two samples, D^2 is defined by the equation

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

in which the allowance for bias $(1/\lambda^2)$ included in the unstudentized form has been dropped. Such adjustments are, of course, unnecessary when the correct sampling distribution is available. This point deserves emphasis, since some statisticians, unfamiliar with the use of exact distributions, still seem to regard the discussion of bias as relevant to problems of estimation.

In a very brilliant research R. C. Bose & S. N. Roy have demonstrated that the distribution of D^2, so defined, takes a form derivable from distribution (C) of my 1928 paper, of which the frequency element is

$$\frac{\frac{1}{2}(n-1)!}{\frac{1}{2}(p-2)!\,\frac{1}{2}(n-p-1)!} R^{p-2}(1-R^2)^{\frac{1}{2}(n-p-1)} e^{-\frac{1}{2}\beta^2}\left\{1+\frac{n+1}{1.p}\frac{R^2\beta^2}{2}+\frac{(n+1)(n+3)}{1.2.p(p+1)}\left(\frac{R^2\beta^2}{2}\right)^2+\dots\right\},$$

which reduces to distribution (B) when $n \to \infty$, and $nR^2 \to B^2$, but for finite n differs from that distribution in replacing the Bessel function by a confluent hypergeometric function.

*Table of 5 % points of the distribution of B**

Values of β	Value of n_1						
	1	2	3	4	5	6	7
0·0	1·9600	2·4477	2·7955	3·0802	3·3272	3·5485	3·7506
0·2	1·9985	2·4720	2·8140	3·0955	3·3405	3·5602	3·7613
0·4	2·1070	2·5419	2·8680	3·1405	3·3796	3·5951	3·7930
0·6	2·2654	2·6497	2·9533	3·2125	3·4426	3·6517	3·8445
0·8	2·4505	2·7855	3·0640	3·3076	3·5268	3·7278	3·9144
1·0	2·6461	2·9398	3·1941	3·4216	3·6291	3·8210	4·0005
1·2	2·8451	3·1059	3·3386	3·5505	3·7462	3·9289	4·1008
1·4	3·0449	3·2796	3·4935	3·6911	3·8756	4·0491	4·2134
1·6	3·2449	3·4584	3·6561	3·8408	4·0148	4·1796	4·3363
1·8	3·4449	3·6410	3·8246	3·9978	4·1620	4·3184	4·4681
2·0	3·6449	3·8263	3·9976	4·1604	4·3158	4·4645	4·6074
2·2	3·8449	4·0137	4·1743	4·3278	4·4750	4·6166	4·7531
2·4	4·0449	4·2027	4·3539	4·4990	4·6388	4·7738	4·9043
2·6	4·2449	4·3932	4·5359	4·6735	4·8065	4·9353	5·0603
2·8	4·4449	4·5847	4·7199	4·8506	4·9774	5·1006	5·2204
3·0	4·6449	4·7772	4·9055	5·0301	5·1512	5·2691	5·3840
3·2	4·8449	4·9705	5·0926	5·2115	5·3273	5·4404	5·5508
3·4	5·0449	5·1644	5·2809	5·3946	5·5056	5·6142	5·7204
3·6	5·2449	5·3589	5·4703	5·5792	5·6857	5·7901	5·8924
3·8	5·4449	5·4914	5·6606	5·7650	5·8675	5·9679	6·0665
4·0	5·6449	5·7493	5·8516	5·9521	6·0506	6·1475	6·2426
4·2	5·8449	5·9451	6·0434	6·1401	6·2351	6·3285	6·4204
4·4	6·0449	6·1412	6·2359	6·3290	6·4206	6·5109	6·5998
4·6	6·2449	6·3376	6·4288	6·5187	6·6072	6·6945	6·7805
4·8	6·4449	6·5342	6·6223	6·7091	6·7947	6·8792	6·9625
5·0	6·6449	6·7311	6·8162	6·9002	6·9831	7·0649	7·1457

* I am indebted to the Council of the Royal Society for permission to reproduce this table, which appeared in *Proc. Roy. Soc.* A, **121**, 665.

The translation of the solution of Bose & Roy into distribution C is now merely

$$\lambda^2 p\Delta^2 = \beta^2,$$
$$\lambda^2 pD^2 = nR^2/(1-R^2) = T^2,$$

and, as in the case of Hotelling's formula,

$$N_1+N_2-2 = n.$$

For large samples it will therefore be usually sufficient to calculate T, or perhaps $T/\sqrt{(1+T^2/n)}$, and to enter the (β, B) table with this value for B; the value of β for which a significant value is just attained will give a fiducially limiting value for $\Delta\lambda\sqrt{p}$.

This procedure brings into relief the desirability of two further extensions of the tables available; (*a*) my table gives only the upper 5 % B for given values of β, the lower values will now also be required; (*b*) it would be most valuable in addition to have tables of distribution C, in the form of T^2 for given β, or some other form which will tend to the available limiting form given by the B table. A few suitably chosen values of n, such as 24, 12, 8, 6, would doubtless suffice to show over what parts of the field the limiting distribution is of sufficient accuracy.

V. Extension of discriminant analysis

We have seen that the Calcutta School have elucidated the notion of generalized distance in fields of multiple variates, and have advanced their researches to a point at which only a moderate extension of existing tables is needed to apply exact tests of significance to this measure. Work on the discriminant functions is not so far advanced. Using the same geometrical analogy, the discriminant function is a unit vector specifying the direction of one population from another. It is true that when two populations are indistinguishable in respect of the measurements available, no significant estimate can be made either of the distance or of the direction. Hotelling's insight thus led him to the appropriate basic test of significance for both problems. When, however, Hotelling's test of significance is satisfied, the relevant problems which suggest themselves diverge. Measuring distance, we naturally will ask whether one observed distance significantly exceeds another. Measuring direction, we shall likewise be led to test whether three or more populations are collinear, or coplanar. The relevance, or even urgency, of such questions in all fields in which populations are discriminated by multiple measurements is obvious.

Let us suppose we have s populations designated by $\pi = 1, \ldots, s$ represented by samples of N^π individuals. The means of the p variates in sample π will be $\bar{x}_{\pi i}$, $i = 1, \ldots, p$.

Any component of the set of possible comparisons among samples may be defined by a set of values λ^π such that

$$\sum_{\pi=1}^{s} \frac{\lambda^{\pi 2}}{N^\pi} = 1;$$

we may speak of different comparisons as orthogonal if they are specified by λ and μ satisfying the condition

$$\sum_{\pi=1}^{s} \frac{\lambda^\pi \mu^\pi}{N^\pi} = 0.$$

If

$$\sum_{\pi=1}^{s} N^{\pi} = N,$$

we may choose as our first component

$$\kappa^{\pi} = N^{\pi}/\sqrt{N}.$$

Then κ will not represent a comparison among populations, but by virtue of the orthogonal property its inclusion ensures that

$$\sum_{\pi=1}^{s} \lambda^{\pi} = 0,$$

so that all components after the first will, properly speaking, be comparisons among populations. These comparisons are analogous to the second set of variates $x_1 \ldots x_s$ discussed by Hotelling (1936). Hotelling, however, is considering a set of normal variates, whereas λ^{π}/N^{π}, which may be regarded as a variate, varying from sample to sample, need make no approach to a normal distribution.

For any such comparison we may now put

$$d_i = \sum_{\pi=1}^{s} \lambda^{\pi} \bar{x}_{\pi i},$$

and obtain the corresponding discriminant function

$$X = \sum_{i=1}^{p} b^i x_i,$$

by using the relations

$$b^i = \sum_{j=1}^{p} s^{ij} d_j = \sum_{j=1}^{p} \sum_{\pi=1}^{s} s^{ij} \bar{x}_{\pi j} \lambda^{\pi}$$

$$= \sum_{\pi=1}^{s} \lambda^{\pi} \xi^i_{\pi},$$

where

$$\xi^i_{\pi} = \sum_{j=1}^{p} s^{ij} \bar{x}_{\pi j}.$$

In the analysis above, s^{ij} has its former meaning as an element of the information matrix within samples; it is seen that ξ^i_{π} is independent of the comparison chosen. A set of $(s-1)$ functionally independent comparisons thus gives a set of $(s-1)$ vectors defining the general discriminant space appropriate to the s samples.

If now

$$\sum_{i=1}^{p} \bar{x}_{i\chi} \xi^i_{\pi} = t_{\pi\chi},$$

the sum of squares among populations for any chosen comparison is

$$\sum_{i=1}^{p} d_i b^i = \sum_{i=1}^{p} \sum_{j=1}^{p} s^{ij} d_i d_j$$

$$= \sum_{\pi=1}^{s} \sum_{\chi=1}^{s} \lambda^{\pi} \lambda^{\chi} \sum_{i=1}^{p} \sum_{j=1}^{p} s^{ij} \bar{x}_{\pi i} \bar{x}_{\chi j}$$

$$= \sum_{\pi=1}^{s} \sum_{\chi=1}^{s} \lambda^{\pi} \lambda^{\chi} t_{\pi\chi}.$$

Now the set of values $\lambda^{\pi}/\sqrt{N^{\pi}}$, including κ, may be regarded as defining s mutually orthogonal sets of direction cosines; hence

$$\kappa^{\pi^2}+\lambda^{\pi^2}+\mu^{\pi^2}+\ldots = N^{\pi},$$

$$\kappa^{\pi}\kappa^{\chi}+\lambda^{\pi}\lambda^{\chi}+\mu^{\pi}\mu^{\chi}+\ldots = 0.$$

Adding now the expressions for the sums of squares among populations, the total for a complete orthogonal set of comparisons is found to be

$$\sum_{\pi=1}^{s} N^{\pi}t_{\pi\pi};$$

but the component κ gives

$$\sum_{\pi=1}^{s}\sum_{\chi=1}^{s}\kappa^{\pi}\kappa^{\chi}t_{\pi\chi}=\frac{1}{N}\sum_{\pi=1}^{s}\sum_{\chi=1}^{s}N^{\pi}N^{\chi}t_{\pi\chi},$$

so that the remaining $s-1$ comparisons contain together

$$\sum_{\pi=1}^{s}N^{\pi}t_{\pi\pi}-\frac{1}{N}\sum_{\pi=1}^{s}\sum_{\chi=1}^{s}N^{\pi}N^{\chi}t_{\pi\chi}.$$

It follows that no component can be chosen so that

$$\sum_{\pi=1}^{s}\sum_{\chi=1}^{s}\lambda^{\pi}\lambda^{\chi}t_{\pi\chi}$$

exceeds this amount.

* Each comparison has p degrees of freedom, so that in testing for collinearity we may maximize the sum of squares of X among samples, deduct this amount from the total, and test whether the remaining $(s-2)p$ degrees of freedom contain a larger sum of squares than the variation within samples will account for. Likewise coplanarity will be tested by deducting the largest pair of mutually orthogonal components.

VI. Significant differences in direction

If this procedure is applied to three samples, two orthogonal comparisons are found, containing respectively the maximal and the minimal sums of squares among samples. These may be conveniently found from the symmetric functions

$$\theta_1+\theta_2 = N^1t_{11}+N^2t_{22}+N^3t_{33}-\frac{1}{N}(N^{1^2}t_{11}+N^{2^2}t_{22}+N^{3^2}t_{33})$$

and

$$\theta_1\theta_2 = \frac{N^1N^2N^3}{N}\{t_{22}t_{33}-t_{23}^2-2t_{11}t_{23}+2t_{12}t_{13}$$
$$+t_{33}t_{11}-t_{31}^2-2t_{22}t_{31}+2t_{12}t_{23}$$
$$+t_{11}t_{22}-t_{12}^2-2t_{33}t_{12}+2t_{13}t_{23}\}.$$

* For author's revised version of paper from this point on, see end of paper (page 142).

The question whether the contrast in the available measurements between two given samples differs in direction from that supplied by a hypothetical discriminant function

$$\Xi = \sum_{i=1}^{p} \beta^i x_i$$

may be solved by considering the limiting case in which a third sample is introduced, very distant from the other two in the required direction, i.e. if

$$\delta_i = \sum_{j=1}^{p} \beta^i s_{ij},$$

we let

$$\bar{x}_{3i} = \Omega\delta_i,$$

where Ω shall be increased without limit. The question whether the discriminant function observed for the two given populations differs significantly from Ξ may be recognized to be the same as the question whether the three populations depart significantly from collinearity.

For resolving this question, we have

$$t_{13} = \sum_{i=1}^{p}\sum_{j=1}^{p} s^{ij}\bar{x}_{1i}\bar{x}_{3j}$$

$$= \Omega \sum_{i=1}^{p} \bar{x}_{1i} \sum_{j=1}^{p} s^{ij}\delta_j$$

$$= \Omega \sum_{i=1}^{p} \beta^i \bar{x}_{1i}.$$

Similarly,

$$t_{23} = \Omega \sum_{i=1}^{p} \beta^i \bar{x}_{2i},$$

and finally

$$t_{33} = \Omega^2 \sum_{i=1}^{p}\sum_{j=1}^{p} \beta^i\beta^j s_{ij};$$

since the absolute magnitude of the coefficient β is so far arbitrary, we may equate t_{33} to Ω^2 by adopting the convention that

$$\sum_{i=1}^{p}\sum_{j=1}^{p} \beta^i\beta^j s_{ij} = 1.$$

As Ω is increased, the larger root of the quadratic, θ_1, also increases indefinitely. Significance of departure from collinearity is tested by the smaller root. The limiting value of this smaller root may be conveniently found from the ratio $\theta_1\theta_2/(\theta_1+\theta_2)$. The numerator and the denominator of this each contains a portion proportional to Ω^2.

Thus,

$$(\theta_1+\theta_2)/\Omega^2 \to \frac{N^3(N-N^3)}{N\Omega^2} t_{33}$$

$$= \frac{N^3(N-N^3)}{N},$$

and

$$\theta_1\theta_2/\Omega^2 \to \frac{N^1N^2N^3}{N\Omega^2}\{(t_{11}-2t_{12}+t_{22})\,t_{33}-(t_{13}^2-2t_{13}t_{23}+t_{23}^2)\}$$

$$= \frac{N^1N^2N^3}{N}\left\{\sum_{i=1}^{p}\sum_{j=1}^{p} s^{ij}(\bar{x}_{1i}-\bar{x}_{2i})(\bar{x}_{1j}-\bar{x}_{2j}) - \sum_{i=1}^{p}{}^2\, \beta^i(\bar{x}_{1i}-\bar{x}_{2i})\right\}.$$

Hence,
$$\theta_2 \to \frac{N^1 N^2}{N - N^3} \left\{ \sum_{i=1}^{p} \sum_{j=1}^{p} s^{ij} d_i d_j - (\Xi_1 - \Xi_2)^2 \right\},$$

where, as in the previous work with two samples, d stands for the difference between their means in any measurement.

Replacing $N - N^3$ by $N^1 + N^2$, we observe that the number N_3 has disappeared, as irrelevant to the analysis; the factor outside the bracket is now λ^2 of our previous sections. The first term is recognizable as T^2, and the whole may be written

$$T^2 - \lambda^2(\Xi_1 - \Xi_2)^2,$$

which simply replaces T^2 in Hotelling's test of significance.

In the form of the analysis of variance we have

Degrees of freedom	Sum of squares
$p-1$	$T^2 - \lambda^2(\Xi_1 - \Xi_2)^2$
$n-p+1$	n

The significance of the deviation of the observed discriminant function from any proposed function of the same kind is thus easily tested. The test may, however, be thrown into another form of great simplicity.

If $b^1, \ldots, b^p$ are the coefficients of the discriminant function obtained from the two samples by the regression method, we have seen that

$$n \sum_{j=1}^{p} s_{ij} b^j = \lambda^2 d_i \left(1 - \sum_{j=1}^{p} b^j d_j\right).$$

Hence
$$n \sum_{i=1}^{p} \sum_{j=1}^{p} s_{ij} b^i b^j = \lambda^2 \sum_{i=1}^{p} b^i d_i \left(1 - \sum_{i=1}^{p} b^i d_i\right)$$
$$= \lambda^2 R^2 (1 - R^2).$$

Also
$$n \sum_{i=1}^{p} \sum_{j=1}^{p} s_{ij} b^j \beta^i = \lambda^2 (1 - R^2) \sum_{i=1}^{p} \beta^i \delta_i$$
$$= \lambda^2 (1 - R^2)(\Xi_1 - \Xi_2).$$

Hence if r is the correlation coefficient *within samples* between the discriminant function proposed, and that obtained,

$$r^2 = \left(\sum_{i=1}^{p} \sum_{j=1}^{p} s_{ij} b^j \beta^i\right)^2 \Big/ \left(\sum_{i=1}^{p} \sum_{j=1}^{p} s_{ij} b^i b^j\right)\left(\sum_{i=1}^{p} \sum_{j=1}^{p} s_{ij} \beta^i \beta^j\right)$$
$$= \frac{\lambda^2}{n} \frac{1 - R^2}{R^2} (\Xi_1 - \Xi_2)^2.$$

Now, we have already shown that

$$\frac{T^2}{n} = \frac{R^2}{1 - R^2},$$

so the expression for the smaller root may be written simply as

$$T^2(1 - r^2).$$

We should thus reject any proposed discriminant function, if its correlation r within samples with the best discriminant obtainable is so low that the variance ratio

$$e^{2z} = \frac{T'^2(1-r^2)}{p-1} \cdot \frac{n-p+1}{n}$$

is significant for $n_1 = p-1$, $n_2 = n-p+1$; T^2 being Hotelling's generalization of Student's ratio.

Example. Four measurements on the flowers of fifty plants each from the species *Iris versicolor* and *I. setosa* (Fisher, 1936, p. 183) give a ratio of the sums of squares 26·335, for 4 against 95 degrees of freedom. The 1 % z (3, 95) is 0·6926, and the corresponding variance ratio is 3·995. The ratio of the sums of squares is therefore

$$\frac{3}{95} \times 3{\cdot}995 = {\cdot}1261.$$

This is only 0·004791 of that observed. Consequently, at a 1 % level of significance we should reject formulae having a correlation with that obtained less than $\sqrt{0{\cdot}995209}$, or 0·99760. This is a convenient and direct measure of the precision of the discriminant function as estimated.

VII. Summary

The results of three independent lines of research on the treatment of multiple measurements are set out in a consistent notation. The method is extended to the examination of collinearity and coplanarity of samples, and to testing the significance of deviations in direction.

REFERENCES

M. M. Barnard (1935). "The secular variations of skull characters in four series of Egyptian skulls." *Ann. Eugen., Lond.*, **6**, 352–71.

R. C. Bose (1936). "On the exact distribution of D^2 statistic." *Sankhya*, **2**, 143–54.

R. C. Bose & S. N. Roy (1938). "The exact distribution of the Studentized D^2 statistic." *Sankhya*, **3**, pt. 4.

H. Fairfield Smith (1936). "A discriminant function for plant selection." *Ann. Eugen., Lond.*, **7**, 240–50.

R. A. Fisher (1928). "The general sampling distribution of the multiple correlation coefficient." *Proc. Roy. Soc.* A, **121**, 654–73.

—— (1936). "The use of multiple measurements in taxonomic problems." *Ann. Eugen., Lond.*, **7**, 179–88.

H. Hotelling (1931). "The generalization of 'Student's' ratio." *Ann. Math. Statist.* **2**, 360–78.

—— (1936). Relation between two sets of variates. *Biometrika* **28**, 321–77.

P. C. Mahalanobis (1927). "Analysis of race mixture in Bengal." *J. Asiat. Soc. Beng.* **23**, 301–33.

—— (1930). "On tests and measures of group divergence. Part I. Theoretical formulae." *J. Asiat. Soc. Beng.* **26**, 541–88.

—— (1936). "On the generalized distance in statistics." *Proc. Nat. Inst. Sci. Ind.* **12**, 49–55.

N. Wallace & R. M. W. Travers (1938). "A psychometric sociological study of a group of speciality salesmen." *Ann. Eugen., Lond.*, **8**, 266–302.

Author's revised version of last section of paper (CMS 33.385a) - see page 383.

Each comparison has p degrees of freedom, so that in testing for collinearity we may maximize the sum of squares of X among samples, deduct this amount from the total, and test whether the remaining degrees of freedom contain a larger sum of squares than the variation within samples will account for. Likewise coplanarity will be tested by deducting the largest pair of mutually orthogonal components.

In choosing the component making the largest contribution it will be observed that $s - 2$ coefficients have been adjusted; consequently the number of degrees of freedom to be ascribed to the first component is $p + s - 2$, leaving $(s - 2)(p - 1)$ for the remainder. Similarly, if the next largest component be separated, it will contain $p + s - 4$ and leave $(s - 3)(p - 2)$. It will be noticed that the sum of the arithmetic series $(p + s - 2) + (p + s - 4) + \cdots + (p - s + 2)$ adds to the total of $p(s - 1)$ degrees of freedom for the $(s - 1)$ components and represents the partition of degrees of freedom among them if they are chosen in order of magnitude.

VI. Significant differences in direction

The question whether the contrast in the available measurements between two given samples differs significantly in direction from that supplied by a hypothetical discriminant function

$$X' = \Sigma\beta^i x_i$$

may be most simply resolved by eliminating the variate X', or by using the partial variation only for all variates when X' is fixed.

Thus, in the regression problem of Section I, the sum of the products with y becomes

$$S(X'y) = \lambda^2\Sigma\beta^i d_i = \lambda^2 D$$

where D is the difference between the mean values of X' in the two samples.

The sums of squares of X' within samples may be simply expressed in terms of the correlation *within samples* between X' and X, the discriminant function obtained from the observations; for the sum of squares within samples of X is

$$\lambda^2 R^2(1 - R^2)$$

and the sum of the products XX' is

$$\lambda^2 D(1 - R^2)$$

whence, if r is the correlation of X with X', it follows that the sum of the squares of X' is

$$\lambda^2 D^2(1 - R^2)/R^2 r^2:$$

whence adding $\lambda^2 D$ that for all observations must be

$$\lambda^2 D^2 \{1 - R^2(1 - r^2)\}/R^2 r^2.$$

We may see that thus the elimination of X' reduces the sum of squares for y from λ^2 to

$$\lambda^2 - (\lambda^2 D)^2 \cdot R^2 r^2/\lambda^2 D^2(1 - R^2 + r^2 R^2)$$

$$= \lambda^2 \left(1 - \frac{R^2 r^2}{1 - R^2 + r^2 R^2}\right) = \frac{\lambda^2(1 - R^2)}{1 - R^2 + r^2 R^2},$$

while the portion expressible in terms of the variates x is reduced from $\lambda^2 R^2$ to

$$\lambda^2 R^2 \left(1 - \frac{r^2}{1 - R^2 + R^2 r^2}\right) = \frac{\lambda^2 R^2(1 - R^2)(1 - r^2)}{1 - R^2 + r^2 R^2}.$$

The ratio of the part to a whole has thus been changed from R^2 to $R^2(1 - r^2)$. If when so reduced the multiple correlation is no longer significant, then the hypothetical discriminant function X' is not contradicted by the data. The whole class of discriminant functions contradicted by the data at any chosen level of significance is thus specified simply by the correlation coefficient within samples between the function proposed, and that calculated from the data themselves.

Example. Four measurements on the flowers of fifty plants each from the species *Iris versicolor* and *I. setosa* (Fisher, 1936, p. 180) give a ratio of the sums of squares

$$R^2 = \cdot 963416,$$

for 4 against 95 degrees of freedom. The 1% $z(3,95)$ is ·6926, and the corresponding variance ratio is 3·995. Here R^2 is

$$\frac{3 \times 3\cdot 995}{95 + 3 \times 3\cdot 995} = \cdot 112025.$$

The ratio of R^2 required for significance at the 1% level, to the value of R^2 observed, is

$$1 - r^2 = \cdot 116279$$

where

$$r = \cdot 94006$$

given the minimal value of the correlation within populations, such that any discriminant function proposed having a lower correlation than this may be rejected at the 1% level. This is a convenient and direct measure of the precision of the discriminant function as estimated.