THE STATISTICAL SIGNIFICANCE OF CANONICAL CORRELATIONS

BY M. S. BARTLETT

1. In an important paper published in this *Journal*, Hotelling (1936) has shown that the generalized variance matrix*

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix}$$

of a vector variate $\mathbf{x}$ which has been partitioned into two parts $\mathbf{x}_1$ and $\mathbf{x}_2$ with, say, q and p components, can, by appropriate linear transformations $\mathbf{L}_1\mathbf{x}_1$ and $\mathbf{L}_2\mathbf{x}_2$ of $\mathbf{x}_1$ and $\mathbf{x}_2$, be thrown into the canonical form

$$\begin{pmatrix} \mathbf{L}_1\mathbf{V}_{11}\mathbf{L}_1' & \mathbf{L}_1\mathbf{V}_{12}\mathbf{L}_2' \\ \mathbf{L}_2\mathbf{V}_{21}\mathbf{L}_1' & \mathbf{L}_2\mathbf{V}_{22}\mathbf{L}_2' \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{R} \\ \mathbf{R} & 1 \end{pmatrix}.$$

$\mathbf{R}$ is a rectangular matrix which is zero except for a leading diagonal of squares λ_i^2 of canonical correlations.

Similar operations on the estimated matrix variance V give rise to estimated canonical correlations l_i, which measure the correlations between estimates of the linear functions $\mathbf{L}_1\mathbf{x}_1$ and $\mathbf{L}_2\mathbf{x}_2$. While Hotelling has given asymptotic standard errors for the coefficients l_i, it is known that the significance of these correlations, as in the simple case $p = 1$, is more generally to be interpreted as the significance of the regression relations of $\mathbf{x}_2$ with $\mathbf{x}_1$, the validity of any exact tests of significance depending on the supposition that the dependent variate $\mathbf{x}_2$, apart from its linear dependence on $\mathbf{x}_1$, is normal.

Special cases of the simultaneous distribution of the correlations l_i, when $\mathbf{x}_2$ and $\mathbf{x}_1$ are unrelated, have been considered by Hotelling (1936) and Girschick (1939), but an important theoretical advance is represented by the derivation of the distribution (under the same conditions) for any values of p and q (Fisher, 1939; Hsu, 1939). It will be shown that this distribution makes available further possible tests; and since the problem of the most appropriate tests of significance

* A matrix is usually denoted by a capital letter, and if it has both a population and sample value, the population value is given in heavier type (cf. Bartlett, 1939). The transpose of any matrix A is denoted by A'. A matrix with only one column is a vector, and is often denoted by a small letter. To avoid confusion, a vector variate $\mathbf{x}$ is written in heavier type throughout, to distinguish it from a single variate x. If $\mathbf{x}$ is measured from its population mean, the variance matrix $\mathbf{V}$ is the average value of $\mathbf{x}\mathbf{x}'$. In practice we lose one or more degrees of freedom by measuring $\mathbf{x}$ from sample or regression means, but without loss of generality we shall suppose that our sample consists of measurements of $\mathbf{x}$ with ν degrees of freedom.

has not always been considered very adequately by other writers, it is also the purpose of this paper to explain the logical relation of these further tests to tests of significance previously available.*

2. For the case $p \leqslant q$, the distribution of l_i, when these roots are arranged in order of magnitude, is given by

$$F(l_1^2, l_2^2, \ldots, l_p^2)\, dl_1^2 dl_2^2 \ldots dl_p^2,$$

where
$$F = C \prod_{i=1}^{p} \left\{ (l_i^2)^{\frac{1}{2}(q-p-1)} (1-l_i^2)^{\frac{1}{2}(\nu-q-p-1)} \prod_{j=i+1}^{p} (l_i^2 - l_j^2) \right\}$$

and
$$C = \pi^{\frac{1}{2}p} \prod_{i=1}^{p} \frac{\Gamma\frac{1}{2}(\nu-i+1)}{\Gamma\frac{1}{2}(q-i+1)\,\Gamma\frac{1}{2}(\nu-q-i+1)\,\Gamma\frac{1}{2}(p-i+1)}. \qquad (1)$$

For $p > q$, we need only reverse the roles of $\mathbf{x}_1$ and $\mathbf{x}_2$.

A criterion which is useful in detecting the simultaneous departure of several roots λ_i from zero is the product

$$\prod_{i=1}^{p} (1-l_i^2) = \Lambda, \quad \text{say.}\dagger$$

When $p = 1$, the distribution of Λ is equivalent to that of l_1^2, and the distribution in (1) can be transformed if required into Fisher's z-distribution. When $p = 2$, it was found by Wilks that a similar distribution exists for $\sqrt{\Lambda}$. For $p > 2$, no exact test is at present available, but the formula

$$\chi^2 = -\{\nu - \tfrac{1}{2}(p+q+1)\} \log \Lambda,$$

with pq degrees of freedom, gives a good approximate test (Bartlett, 1938).

If the roots $\lambda_2^2, \ldots, \lambda_p^2$ are zero, we are, however, including in Λ irrelevant degrees of freedom which might possibly obscure the significance of λ_1^2. For any test on λ_1^2 by itself, we have little choice but to consider l_1^2, though we do not really know whether l_1^2 is the root corresponding to λ_1^2 or not. The probability distribution‡ $p(l_1^2)$ is theoretically obtainable from (1), and hence also 0·05 or 0·01 levels

* The distribution of l_i^2 obtained by Fisher and Hsu has also been obtained by Roy (1939), though this writer was concerned with the different problem of comparing the dispersion in two multivariate normal samples. For a single variate, testing the significance of a sum of squares separated off from the total sum of squares by a multiple correlation or regression formula is equivalent to testing the ratio of two variances, a criterion also employed to test the equality of two population variances. Roy has proposed generalizing the latter problem along lines which give rise to the same distribution problem solved by Fisher and Hsu, but while he has independently obtained the same general distribution, the need for some care in the choice of tests in multivariate analysis is even more evident in the problem with which Roy was concerned. It is obvious, for example, that the p roots which Roy considered cannot represent *all* the possible differences among the $\frac{1}{2}p(p+1)$ variance and covariance parameters between two p-variate normal samples, and some explanation of their interpretation seems required.

† This criterion has been proposed by Wilks (1932), Bartlett (1934), and Hotelling (1936), the last-named denoting it by z.

‡ The probability of a random variable having a particular value x is denoted by $p(x)$. If the variable has a continuous range of values, $p(x)$ denotes the probability of the variable falling in the interval $x, x+dx$. The corresponding notations $x \mid y$ and $p(x \mid y)$ are used when the variable is only being considered for a fixed value y of another variable. The probability symbol p is not of course to be confused with the number p of components in the vector variate $\mathbf{x}_2$.

of significance of l_1^2 for specified values of ν, p and q. The tabulation of these levels would be useful, but would also be a task of some magnitude, and it is therefore worth noting that owing to the problem of identification, the largest root l_1^2 is not a sufficient statistic for λ_1^2, and $p(l_1^2)$ has no unique relevance. If we consider, instead, the distribution of l_1^2 for given values of $l_2^2 \dots l_p^2$, we have corresponding to the probability relation

$$p(l_1^2, \dots, l_p^2) = p(l_1^2 \mid l_2^2, \dots, l_p^2)\, p\,(l_2^2, \dots, l_p^2),$$

the probability density relation

$$F(l_1^2, \dots, l_p^2) = f_1(l_1^2 \mid l_2^2, \dots, l_p^2) f_2(l_2^2, \dots, l_p^2),$$

where the function f_1, apart from the constant term f_2, is determined at once from the function F.

In the logical situation we are postulating where λ_1^2, but not the other roots, is different from zero, it is not evident which distribution, $p(l_1^2)$ or $p(l_1^2 \mid l_2^2, \dots, l_p^2)$, provides the more powerful test, owing to the absence of sufficiency properties, and it is of some interest to consider in detail another problem which is trivial in itself, but serves to illustrate the principles involved.

3. Suppose we have a pair of variates x_1 and x_2 both independently following a rectangular distribution

$$p\,(x) = dx, \quad (0 \leqslant x \leqslant 1).$$

One variate (unspecified) is then shifted a distance α, so that it follows the distribution

$$p(x) = dx, \quad (\alpha \leqslant x \leqslant 1 + \alpha).$$

If x_1 and x_2 denote the variates in order of magnitude, we shall detect the shift α from the larger value, x_1, if α is large enough. To compare the value of $p(x_1)$ and $p(x_1 \mid x_2)$, we note first of all that when $\alpha = 0$,

$$p(x_1) = 2x_1\, dx_1, \quad (0 \leqslant x_1 \leqslant 1)$$

$$p(x_1 \mid x_2) = \frac{dx_1}{1 - x_2}. \quad (x_2 \leqslant x_1 \leqslant 1)$$

For the significance level ϵ, $p(x_1)$ gives a critical value $x_1 = \sqrt{(1-\epsilon)}$, while $p(x_1 \mid x_2)$ gives $x_1 = 1 - \epsilon(1 - x_2)$. If α is different from zero, a peculiar feature (analogous to the canonical correlation problem) is that the larger observation x_1 may or may not be associated with α. For $p(x_1)$ we find

$$(2x_1 - \alpha)\, dx_1, \quad (\alpha \leqslant x_1 \leqslant 1)$$

$$dx_1. \quad (1 < x_1 \leqslant 1 + \alpha)$$

For $p(x_1 \mid x_2)$, we have

$$\frac{2dx_1}{\alpha + 2(1 - x_2)}, \quad (x_2 \leqslant x_1 \leqslant 1)$$

$$\frac{dx_1}{\alpha + 2(1 - x_2)}. \quad (1 < x_1 \leqslant 1 + \alpha)$$

This is provided $x_2 \geqslant \alpha$; for $x_2 < \alpha$, we have

$$p(x_1 \mid x_2) = dx_1. \quad (\alpha \leqslant x_1 \leqslant 1+\alpha)$$

Using the terminology of Neyman and Pearson, we shall denote the power of the test derived from x_1 by P; and for that from $x_1 \mid x_2$, by P'. Then

$$\begin{aligned} 1-P &= \int_{\alpha}^{\sqrt{(1-\epsilon)}} (2x_1-\alpha)\,dx_1 \\ &= 1-\epsilon-\alpha\sqrt{(1-\epsilon)}. \end{aligned}$$

For $1-P'$, we have first of all, for given x_2, an integral

$$\int_{x_2}^{1-\epsilon(1-x_2)} p(x_1 \mid x_2),$$

which gives

$$1-\epsilon(1-x_2)-\alpha, \quad (x_2<\alpha)$$

$$\frac{2(1-\epsilon)(1-x_2)}{\alpha+2(1-x_2)}. \quad (x_2 \geqslant \alpha)$$

Since $p(x_2 \mid \alpha)$ is given by

$$\begin{aligned} &\{\alpha+2(1-x_2)\}\,dx_2, && (\alpha \leqslant x_2 \leqslant 1) \\ &dx_2, && (0 \leqslant x_2 < \alpha) \end{aligned}$$

we finally obtain, after averaging over x_2,

$$1-P' = (1-\epsilon)(1-\alpha)-\tfrac{1}{2}\alpha^2\epsilon.$$

Before comparing P with P', we may remember that we do not expect either x_1 or $x_1 \mid x_2$ to provide the most powerful test obtainable. Theoretically we can see what this test would be by considering the ratio

$$p(x_1, x_2 \mid \alpha)/p(x_1, x_2 \mid 0) = X_\alpha,$$

say, though since the value of X_α is indeterminate unless the true value of α is specified, it should be realized that X_α does not provide us with any actual test, only with a theoretical upper limit for P or P'.

The criterion X_α has the distribution

$$\begin{array}{rlll} X_\alpha = & \infty & 1 & 0 \\ p(X_\alpha \mid \alpha) = & \alpha & (1-\alpha)^2 & \alpha(1-\alpha) \\ p(X_\alpha \mid 0) = & 0 & (1-\alpha)^2 & 2\alpha-\alpha^2 \end{array}$$

For $(1-\alpha)^2 \geqslant \epsilon$, we shall allow the value $X_\alpha = 1$ to be significant in $\epsilon/(1-\alpha)^2$ of the times that the value 1 occurs; if $(1-\alpha)^2 \leqslant \epsilon$, we allow $X_\alpha = 0$ to be significant in the fraction

$$\frac{\epsilon-(1-\alpha)^2}{2\alpha-\alpha^2}$$

of the times that $X_\alpha = 0$ occurs. The power P'' of a test that could be based on X_α is then

$$\alpha+\epsilon, \quad [(1-\alpha)^2 \geqslant \epsilon],$$

$$\alpha+(1-\alpha)^2+\left\{\frac{\epsilon-(1-\alpha)^2}{2\alpha-\alpha^2}\right\}\alpha(1-\alpha), \quad [(1-\alpha)^2 \leqslant \epsilon].$$

Comparative values of P, P' and P'' are given in Table 1 for $\epsilon = 0{\cdot}05$ and $0{\cdot}10$.

Table 1

ϵ \ α		0	0·1	0·2	0·4	0·6	0·8	0·9
0·05	P	0·0500	0·1475	0·2449	0·4399	0·6348	0·8298	0·9272
	P'	0·0500	0·1453	0·2410	0·4340	0·6290	0·8260	0·9253
	P''	0·0500	0·1500	0·2500	0·4500	0·6500	0·8464	0·9373
0·10	P	0·1000	0·1949	0·2897	0·4795	0·6692	0·8590	0·9538
	P'	0·1000	0·1905	0·2820	0·4680	0·6580	0·8520	0·9505
	P''	0·1000	0·2000	0·3000	0·5000	0·7000	0·8785	0·9603

It will be seen that $p(x_1)$ provides a test in this problem rather more powerful than $p(x_1 \mid x_2)$, but that the latter is quite effective. We cannot of course transfer this result to our main problem, but it is clear that $p(l_1^2 \mid l_2^2, \ldots, l_p^2)$ may justifiably be considered, at least until the distribution $p(l_1^2)$ has been tabulated.

4. Returning then to this distribution, we may examine one or two special cases before formally noting the significance level of l_1^2 in general. It has been shown by Fisher and Hsu that for ν large, the distribution of $l_1^2, l_2^2, \ldots, l_p^2$ tends to

$$G(m_1^2, m_2^2, \ldots, m_p^2)\, dm_1^2 dm_2^2 \ldots dm_p^2,$$

where $m_i^2 = \frac{1}{2}\nu l_i^2$, $\quad G = C' \prod_{i=1}^{p} \left\{ (m_i^2)^{\frac{1}{2}(q-p-1)}\, e^{-m_i^2} \prod_{j=i+1}^{p} (m_i^2 - m_j^2) \right\},$

and
$$1/C' = \prod_{i=1}^{p} \{ \Gamma\tfrac{1}{2}(q-i+1)\, \Gamma\tfrac{1}{2}(p-i+1) \}. \tag{2}$$

For the particular case $p = 2$, $q = 3$, the distribution of $m_1^2 \mid m_2^2$ is

$$(m_1^2 - m_2^2)\, e^{-(m_1^2 - m_2^2)}\, dm_1^2,$$

which is a function simply of $m_1^2 - m_2^2$. If alternatively we consider the distribution $p(m_1^2)$, we obtain
$$2e^{-m_1^2}\{e^{-m_1^2} - (1 - m_1^2)\}\, dm_1^2,$$

the $0{\cdot}05$ significance level for which is $5{\cdot}37$. From $p(m_1^2 \mid m_2^2)$ this value of $5{\cdot}37$ corresponds to a level $0{\cdot}030$ if $m_2^2 = 0$, to $0{\cdot}045$ if m_2^2 is equal to its expected value $0{\cdot}50$, and to $0{\cdot}05$ when m_2^2 reaches the value $0{\cdot}63$. These results merely illustrate how the significance level of m_1^2 depends on which distribution is being used.

For the case $p = 3$, $q = 4$, the significance level for m_1^2 can be written

$$e^{-u}\left\{ (u+1) + \frac{u^2}{2+v} \right\},$$

where $u = m_1^2 - m_2^2$, $v = m_2^2 - m_3^2$. The level of significance thus depends mainly on u, as we should expect, but the effect of v is not negligible. The factor multiplying the exponential varies, for example, when $u = 4$, from $10\frac{1}{3}$ for $v = 1$ to 13 for $v = 0$.

The general expression for the significance level for m_1^2 is

$$\frac{\int_{m_1^2}^{\infty} (m_1^2)^{\frac{1}{2}(q-p-1)}\, e^{-m_1^2} \prod_{i=2}^{p} (m_1^2 - m_i^2)\, dm_1^2}{\int_{m_2^2}^{\infty} (m_1^2)^{\frac{1}{2}(q-p-1)}\, e^{-m_1^2} \prod_{i=2}^{p} (m_1^2 - m_i^2)\, dm_1^2},$$

or, if we write
$$\Gamma_x(\alpha) = \int_x^{\infty} x^{\alpha-1}\, e^{-x}\, dx,$$

by

$$\frac{\Gamma_{m_1^2}(\tfrac{1}{2}[p+q+1]) - \left\{\sum_{i=2}^{p} m_i^2\right\} \Gamma_{m_1^2}(\tfrac{1}{2}[p+q-1]) + \left\{\sum_{i=2}^{p} \sum_{j=i+1}^{p} m_i^2 m_j^2\right\} \Gamma_{m_1^2}(\tfrac{1}{2}[p+q-3])}{\Gamma_{m_2^2}(\tfrac{1}{2}[p+q+1]) - \left\{\sum_{i=2}^{p} m_i^2\right\} \Gamma_{m_2^2}(\tfrac{1}{2}[p+q-1]) + \left\{\sum_{i=2}^{p} \sum_{j=i+2}^{p} m_i^2 m_j^2\right\} \Gamma_{m_2^2}(\tfrac{1}{2}[p+q-3])}. \tag{3}$$

For the more general case of finite ν, we have similarly for l_1^2,

$$\frac{\int_{l_1^2}^{1} (l_1^2)^{\frac{1}{2}(q-p-1)} (1-l_1^2)^{\frac{1}{2}(\nu-q-p-1)} \prod_{i=2}^{p} (l_1^2 - l_i^2)\, dl_1^2}{\int_{l_2^2}^{1} (l_1^2)^{\frac{1}{2}(q-p-1)} (1-l_1^2)^{\frac{1}{2}(\nu-q-p-1)} \prod_{i=2}^{p} (l_1^2 - l_i^2)\, dl_1^2},$$

or, if
$$B_x(\alpha, \beta) = \int_x^1 x^{\alpha-1} (1-x)^{\beta-1}\, dx,$$

by

$$\frac{B_{l_1^2}(\tfrac{1}{2}[p+q+1], \tfrac{1}{2}[\nu-p-q+1]) - \left\{\sum_{i=2}^{p} l_i^2\right\} B_{l_1^2}(\tfrac{1}{2}[p+q-1], \tfrac{1}{2}[\nu-p-q+1]) + \dots}{B_{l_2^2}(\tfrac{1}{2}[p+q+1], \tfrac{1}{2}[\nu-p-q+1]) - \left\{\sum_{i=2}^{p} l_i^2\right\} B_{l_2^2}(\tfrac{1}{2}[p+q-1], \tfrac{1}{2}[\nu-p-q+1]) + \dots}. \tag{4}$$

The dependence of results (3) and (4) not only on ν, p and q, but also on the particular values of $l_2^2, \dots, l_p^2$, makes it impracticable to tabulate the 0·05 or other levels of significance; but it is not difficult in any instance to find the exact level from (3) or (4), using the published tables of $\Gamma_x(\alpha)$ or $B_x(\alpha, \beta)$.*

It must be recognized that if the second root λ_2^2 is also different from zero, the distribution of l_1^2 for given l_2^2 is quite irrelevant, but except possibly when p is rather large, it is probable that two or more non-zero roots would be detected by the Λ criterion, and the testing of λ_1^2 alone by means of l_1^2 (a test which is still not completely efficient) would not arise.

5. Directly we have established the existence of at least one root λ_1^2, we may always proceed to eliminate this correlation λ_1 and the corresponding pair of canonical variates; and analyse the remainder. The theory of eliminating from $\mathbf{x}_2$ a set of specified variates represented, say, by the vector variate $\mathbf{x}_0$ has been

* *Tables of the Incomplete Γ-function*, ed. K. Pearson (1922, His Majesty's Stationery Office, London); *Tables of the Incomplete Beta-function*, ed. K. Pearson (1934, Biometrika Office, University College, London).

indicated by Bartlett (1939).* As a particular case, $\mathbf{x}_0$ may be a hypothetical set of r canonical variates of $\mathbf{x}_2$, and the criterion $\Lambda(\nu-r, p-r, q)$ for the remaining variate $\mathbf{x}_{2\cdot 0}$, in place of the original criterion $\Lambda(\nu, p, q)$ for $\mathbf{x}_2$, would test the goodness of fit of the hypothetical vector canonical variate $\mathbf{x}_0$. In the case $q = 1$, we have the goodness of fit of a hypothetical discriminant function, the problem of which was first raised by Fisher (1938).

It has, however, also been pointed out (Bartlett, 1938, p. 39) that if the canonical vector variate $\mathbf{x}_0$ has been estimated from the data, the symmetrical relation between $\mathbf{x}_2$ and $\mathbf{x}_1$ will imply that each has only $p-r$ and $q-r$ independent components remaining, the χ^2 approximation for the criterion

$$\Lambda' = \prod_{i=r+1}^{p} (1-l_i^2)$$

being $$-\{(\nu-r)-\tfrac{1}{2}[(p-r)+(q-r)+1]\}\log\Lambda' = -\{\nu-\tfrac{1}{2}(p+q+1)\}\log\Lambda',$$

with $(p-r)(q-r)$ degrees of freedom. It was stressed that this reduction of the degrees of freedom essentially depends on the existence of non-zero roots $\lambda_1^2, \ldots, \lambda_r^2$, so that the vector variate $\mathbf{x}_0$ is well-determined, and any effect of selection of $l_1^2, \ldots, l_r^2$ from $l_1^2, \ldots, l_p^2$ can be neglected. Under the same conditions, we may approximately use the tests known for $p = 1$ or 2, for the criterion $\Lambda'(\nu-r, p-r, q-r)$, when $p-r = 1$ or 2.

6. To demonstrate the reduction in degrees of freedom in the case $r = 1$, consider the case when ν is large, and

$$-\nu\log\Lambda \rightarrow \sum_{i=1}^{p} \nu l_i^2 \rightarrow \chi^2.$$

If $\nu l_i^2 = \theta_i$, the determinantal equation for θ_i is of the form

$$|A-\theta\mathbf{V}| = 0,$$

where $\mathbf{V}$ denotes the variance matrix of $\mathbf{x}_2$, and A is a matrix of the sums of squares and products among the p variates of $\mathbf{x}_2$ for that portion of the sample separated off in terms of the independent vector variate $\mathbf{x}_1$. Without loss of generality, we shall suppose that $\mathbf{V} = 1$.

Regarding the ν observations for any variate as a vector with ν orthogonal components, let us now add to the chance variation of the first variate of $\mathbf{x}_2$ a part dependent on each of the q (orthogonal) variates of $\mathbf{x}_1$. For each variate of $\mathbf{x}_1$, the length of the vector representing the first variate of $\mathbf{x}_2$ will then receive an addition X_k, say, $(k = 1 \ldots q)$, which will be of order $\sqrt{\nu}$. Partitioning off the first variate of $\mathbf{x}_2$, we obtain, as our new equation for θ,

$$\left|\begin{array}{c:c} a_{11}+2\Sigma x_1 X+\Sigma X^2-\theta & a_{1j}+\Sigma x_j X \\ \hdashline a_{i1}+\Sigma x_i X & a_{ij}-\theta \end{array}\right| = 0.$$

* See equation (2.8) of the paper cited, and the immediately preceding equation.

The summation sign is for the q degrees of freedom of $\mathbf{x}_1$, and $a_{ij} = \Sigma x_i x_j$, where $x_1, \ldots, x_p$ are the p variates of $\mathbf{x}_2$. Solving the equation for the largest root, we have

$$\theta_1 = \Sigma X^2 + 2\Sigma x_1 X + a_{11} + \sum_{i=2}^{p} \frac{(\Sigma x_i X)^2}{\Sigma X^2} + o\left(\frac{1}{\sqrt{(\Sigma X^2)}}\right).$$

If we neglect the last term, the sum of the remaining roots becomes

$$\sum_{i=1}^{p} \theta_i - \theta_1 = \sum_{i=2}^{p} \left\{a_{ii} - \frac{(\Sigma x_i X)^2}{\Sigma X^2}\right\},$$

which is a χ^2 with $(p-1)(q-1)$ degrees of freedom.

7. To illustrate the use of this test we may consider the data from Kelley quoted by Hotelling (1936), these consisting of correlations among tests in (1) reading speed, (2) reading power, (3) arithmetic speed and (4) arithmetic power, the sample being one of 140 seventh-grade school children. Hotelling, investigating the relation of arithmetical with reading abilities, found canonical correlations

$$l_1 = 0{\cdot}3945, \quad l_2 = 0{\cdot}0688.$$

Since $\nu = 139$, $p = 2$, $q = 2$, the first correlation gives a contribution to χ^2 of

$$-\{139 - \tfrac{1}{2}(2+2+1)\}\log(1 - 0{\cdot}3945^2) = 23{\cdot}09.$$

Similarly the contribution from l_2 is 0·64. The χ^2 analysis is consequently summarized as in Table 2.

Table 2

	D.F.	χ^2
l_1	3	23·09
l_2	1	0·64
Total	4	23·73

It is evident at once, as Hotelling concluded from other tests, that there is a significant relation between arithmetical and reading abilities, which arises entirely from the first canonical correlation.

REFERENCES

Bartlett, M. S. (1934). "The vector representation of a sample." *Proc. Camb. Phil. Soc.* **30**, 327–40.

—— (1938). "Further aspects of the theory of multiple regression." *Proc. Camb. Phil. Soc.* **34**, 33–40.

—— (1939). "A note on tests of significance in multivariate analysis." *Proc. Camb. Phil. Soc.* **35**, 180–5.

Fisher, R. A. (1938). "The statistical utilization of multiple measurements." *Ann. Eugen.* **8**, 376–86.

—— (1939). "The sampling distribution of some statistics obtained from non-linear equations." *Ann. Eugen.* **9**, 238–49.

Girschick, M. A. (1939). "On the sampling theory of the roots of determinantal equations." *Ann. Math. Statist.* **10**, 203–24.

Hotelling, H. (1936). "Relations between two sets of variates." *Biometrika*, **28**, 321–77.

Hsu, P. L. (1939). "On the distribution of roots of certain determinantal equations." *Ann. Eugen.* **9**, 250–8.

Roy, S. N. (1939). "p-statistics or some generalizations in analysis of variance appropriate to multivariate problems." *Sankhyā*, **4**, 381–96.

Wilks, S. S. (1932). "Certain generalizations in the analysis of variance." *Biometrika*, **24**, 471–94.