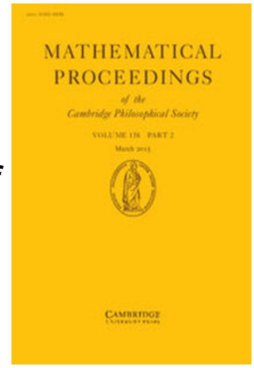

A note on tests of significance in multivariate analysis

M. S. Bartlett

Mathematical Proceedings of the Cambridge Philosophical Society / Volume 35 / Issue 02 / April 1939, pp 180 - 185
DOI: 10.1017/S0305004100020880, Published online: 24 October 2008

Link to this article: http://journals.cambridge.org/abstract_S0305004100020880

How to cite this article:
M. S. Bartlett (1939). A note on tests of significance in multivariate analysis. Mathematical Proceedings of the Cambridge Philosophical Society, 35, pp 180-185 doi:10.1017/S0305004100020880

A NOTE ON TESTS OF SIGNIFICANCE IN MULTIVARIATE ANALYSIS

By M. S. BARTLETT

Received 30 January 1939

1. *Multivariate generalizations.* In multivariate statistical analysis, common terms such as variances and correlation coefficients have received certain generalizations. Wilks (7) has called the determinant $|\mathbf{V}|$, where $\mathbf{V}$ is the matrix* of variances and covariances between several variates, a generalized variance; certain ratios of such determinants have been called by Hotelling (5) vector correlation coefficients and vector alienation coefficients. While these determinantal functions have properties which justify to some extent this kind of generalization, it sometimes seems more reasonable to leave any generalized parameters, or corresponding sample statistics, in the form of matrices of elementary quantities. This is stressed by the formal analogy which then often exists between the generalized and the elementary formulae.

The characteristic function of a vector variate x is the mean value

$$M(t) = E_x(e^{it'x}), \tag{1·1}$$

where x denotes a column vector and t' is a row vector with the same number of elements as x. If x contains $p+q$ component variates which we divide into two subvectors x_1 and x_2 of q and p components respectively, then we may write

$$t'x = (t_1' \mid t_2')\begin{pmatrix} x_1 \\ \hline x_2 \end{pmatrix} = t_1'x_1 + t_2'x_2. \tag{1·2}$$

Correspondingly the matrix $\mathbf{V}$ of variances and covariances is related to the second-degree terms in the expansion of $\log M$ in powers of t by the relation (cf. Aitken (1), equations (3) and (4)),

$$\begin{aligned} t'\mathbf{V}t &= (t_1' \mid t_2')\left(\begin{array}{c|c} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \hline \mathbf{V}_{21} & \mathbf{V}_{22} \end{array}\right)\begin{pmatrix} t_1 \\ \hline t_2 \end{pmatrix} \\ &= t_1'\mathbf{V}_{11}t_1 + 2t_2'\mathbf{V}_{21}t_1 + t_2'\mathbf{V}_{22}t_2. \end{aligned} \tag{1·3}$$

By the generalized variance is meant here the matrix $\mathbf{V}$, not its determinant $|\mathbf{V}|$. The generalized covariance between x_2 and x_1 is the matrix $\mathbf{V}_{21}$ (with trans-

* A matrix of population parameters is denoted by a capital letter of heavier type than the corresponding matrix of sample estimates.

posed matrix $\mathbf{V}'_{21} = \mathbf{V}_{12}$). For the correlation coefficient there is no such obvious generalization; we might, however, define a quantity $\mathbf{R}_{21}$ by the equation

$$\mathbf{R}_{21} = \mathbf{V}_{22}^{-\frac{1}{2}} \mathbf{V}_{21} \mathbf{V}_{11}^{-\frac{1}{2}}. \tag{1·4}$$

Hotelling's analysis (5) of x_2 and x_1 into "canonical components" is obtained by means of the roots of the equation

$$| \mathbf{R}_{21} \mathbf{R}'_{21} - \lambda^2 | = 0, \tag{1·5}$$

where $\mathbf{R}'_{21} = \mathbf{R}_{12}$ is the transpose of $\mathbf{R}_{21}$. An equivalent but more useful equation is

$$| \mathbf{B}_{21} \mathbf{B}_{12} - \lambda^2 | = 0, \tag{1·6}$$

where $\mathbf{B}_{21}$ is the generalized population regression coefficient $\mathbf{V}_{21} \mathbf{V}_{11}^{-1}$, and similarly $\mathbf{B}_{12}$ ($\neq \mathbf{B}'_{21}$) is $\mathbf{V}_{12} \mathbf{V}_{22}^{-1}$. Hotelling's "vector correlation coefficient" $\mathbf{Q}$ and "vector alienation coefficient" $\sqrt{\mathbf{Z}}$ are related to $\mathbf{R}_{21}$ or $\mathbf{B}_{21}$ by the equations

$$\mathbf{Q}^2 = | \mathbf{R}_{21} \mathbf{R}'_{21} | = | \mathbf{B}_{21} \mathbf{B}_{12} |, \tag{1·7}$$

$$\mathbf{Z} = | 1 - \mathbf{R}_{21} \mathbf{R}'_{21} | = | 1 - \mathbf{B}_{21} \mathbf{B}_{12} |. \tag{1·8}$$

Exactly corresponding equations (see Bartlett (3)) hold for sample estimates of these quantities, by the substitution of $C = SS'$ for $\mathbf{V}$, apart from a suitable dividing factor for the estimated matrix of variances and covariances. S denotes the sample matrix of n observations (or deviations) of the vector variate x. The sample quantity corresponding to $\mathbf{Z} = | 1 - \mathbf{B}_{21} \mathbf{B}_{12} |$, defined in general regression terms, is what I have referred to as Λ (Bartlett (2)).

2. *The problem of tests of significance.* If it happens in any problem that we require a general test of the relationship of S_2 with S_1, some single function of these generalized statistics has, of course, to be selected, and it is for this purpose that the criterion Λ, which corresponds to the likelihood criterion, has been used. If we denote below the *sample* regression $C_{21} C_{11}^{-1}$ by B_{21}, the equation for the *sample* canonical correlations l_j can be written in the form

$$| 1 - B_{21} B_{12} - (1 - l^2) | = 0, \tag{2·1}$$

so that

$$\Lambda = | 1 - B_{21} B_{12} | = \prod_{j=1}^{p} (1 - l_j^2). \tag{2·2}$$

The function Λ can potentially detect the departure of any root l_j^2 from zero (it may of course be comparatively insensitive, a weakness shared by any test of this nature). It also has the properties of being invariant not only for linear transformations within x_1 and x_2, but for a reversal of the regression roles of dependent variate (x_2) and independent variate (x_1). Any symmetric function of the invariants l_j^2 which increases (or decreases) for an increase in any l_j^2 for any values of the other l_j^2 has, however, similar properties. Two such functions which we consider are

$$U = \sum_{j=1}^{p} l_j^2, \tag{2·3}$$

and

$$\Omega = \sum_{j=1}^{p} l_j^2/(1-l_j^2). \tag{2·4}$$

The sum U is the trace of the matrix $B_{21}B_{12}$. Since

$$B_{21}B_{12} = C_{21}C_{11}^{-1}C_{12}C_{22}^{-1}, \quad 1-B_{21}B_{12} = C_{22.1}C_{22}^{-1},$$

$$B_{21}B_{12}(1-B_{21}B_{12})^{-1} = C_{21}C_{11}^{-1}C_{12}C_{22.1}^{-1},$$

we have, by writing the equation for l_j^2 as

$$|\, B_{21}B_{12}(1-l^2) - l^2(1-B_{21}B_{12})\,| = 0, \tag{2·5}$$

the result that the sum Ω is the trace of $C_{21}C_{11}^{-1}C_{12}C_{22.1}^{-1}$.

It is important to notice that as test criteria Λ, U and Ω are all equivalent for large n. For if there is no real relationship between S_2 and S_1, we have $l_j^2 = O(n^{-1})$, so that

$$-\log\Lambda = U\{1+O(n^{-1})\} \tag{2·6}$$

$$= \Omega\{1+O(n^{-1})\}. \tag{2·7}$$

The choice of Λ, U or Ω might therefore be considered to some extent as a matter of convenience. To stress their invariant nature, we have expressed them above in terms of the canonical correlations l_j, but certain analyses are possible in terms of the untransformed variates. Thus if

$$S_2 = \begin{pmatrix} S_0 \\ \hdashline S_{2.0} \end{pmatrix}, \quad C_{21} = \begin{pmatrix} C_{01} \\ \hdashline C_{21.0} \end{pmatrix}, \quad C_{22} = \left(\begin{array}{c:c} C_{00} & 0 \\ \hdashline 0 & C_{22.0} \end{array}\right),$$

then

$$\Lambda = \frac{|C_{22.1}|}{|C_{22}|} = \frac{|C_{00.1}|}{|C_{00}|}\frac{|C_{22.01}|}{|C_{22.0}|}. \tag{2·8}$$

Further, U is the trace of

$$C_{01}C_{11}^{-1}C_{10}C_{00}^{-1} + C_{21.0}C_{11}^{-1}C_{12.0}C_{22.0}^{-1}. \tag{2·9}$$

Ω, however, depends on the matrix $C_{21}C_{11}^{-1}C_{12}C_{22.1}^{-1}$; and $C_{22.1}$ is not diagonal like C_{22}.

However, if we let S_1 represent a set of orthogonal vectors in two groups, the factor $C_{21}C_{11}^{-1}C_{12}$ splits into two components. For two vectors only, this factor is of the form $aa' + bb'$, where a and b are column vectors with as many terms as S_2, and the trace of, say, $C_{21}C_{11}^{-1}C_{12}C_{22.1}^{-1}$ is

$$a'C_{22.1}^{-1}a + b'C_{22.1}^{-1}b. \tag{2·10}$$

The additive character of the results (2·9) and (2·10) might appear to give a certain advantage to the criteria U and Ω in comparison with Λ (compare the difference between the χ^2 criterion and the exact likelihood criterion in a simple frequency "goodness of fit" test); but the algebraic analyses above are so far merely formal, and the problem of the corresponding tests of significance has still to be considered. The criterion Ω has been explicitly considered in the form (2·10) by Lawley (6), and also appears implicitly in recent work of Fisher (4), but much more is known of the distribution of Λ. If we write Λ more fully as a function $\Lambda(n, p, q)$, where n is the original number of degrees of freedom in S_2, then $p = 1$ gives the familiar case where Λ, U and Ω are all equivalent to the usual applications of Fisher's z-test. For $p > 1$, the exact distribution of Λ (on the usual assumption of the normality of the dependent vector x_2 for given x_1) is known for q ($q \leqslant p$, say) equal to 1 or 2. Moreover, the characteristic function of $\log \Lambda$ is known in general, an approximate χ^2-test having been obtained from that basis. The two factors in equation (2·8) are also independent*.

3. *Application to discriminant function analysis.* When introducing the important practical problem of discriminating most efficiently between two groups or samples in many variates, Fisher stressed the equivalence of the analysis of S_2 in relation to a single vector S_1 serving to specify the difference between the groups, and the formal multiple regression analysis of S_1 in relation to the variates of S_2. Fisher (4) has extended this formal regression analysis to the case where S_1 is a matrix corresponding, say, to the difference in means of more than two groups. Such an extension is of course covered by the general theory of the relation of S_2 to S_1; in particular, tests of significance will be subject to any of the limitations that have been noted in regard to this general theory.

For $q = 1$, however, exact tests should be available; and we consider below the problem raised by Fisher of testing the difference between an observed and a hypothetical discriminant function. Let $\alpha' S_2 \equiv S_0$ represent the hypothetical discriminant function, and $a' S_2$ the observed discriminant function. Then the first factor in equation (2·8) will test the relation between S_0 and S_1, and the second independent factor, which has the form $\Lambda(n-1, p-1, 1)$, will test the residual relation between S_2 and S_1.

To examine the problem further, we note that

$$U = a_i' C_{22} a_i = (a_i' S_2)^2, \tag{3·1}$$

$$\Omega = a_i' C_{22.1} a_i = (a_i' S_{2.1})^2, \tag{3·2}$$

* For a presentation of these tests in their most general form, see Bartlett (2), p. 338, and (3), p. 38. The exact tests known for $p > 1$ are due to work by Hotelling and Wilks, the case $q = 1$ corresponding to Hotelling's generalization of Student's t-test.

where a_t and a_i are equivalent solutions for a (which has arbitrary absolute magnitude) given by

$$\left.\begin{aligned} a_t &= C_{22}^{-1} C_{21} C_{11}^{-\frac{1}{2}}, \\ a_i &= C_{22.1}^{-1} C_{21} C_{11}^{-\frac{1}{2}}, \end{aligned}\right\} \tag{3·3}$$

(cf. Bartlett(3), p. 37; the suffixes in a_t and a_i denote "total" and "internal", to correspond to the "total variation" represented by C_{22} and the "internal variation" represented by $C_{22.1}$). If we consider the formal analysis of S_2 into S_0 and $S_{2.0}$ given by equation (2·9), we have

$$U = \frac{\alpha' C_{21} C_{11}^{-1} C_{12} \alpha}{\alpha' C_{22} \alpha} + \frac{C_{12.0} C_{22.0}^{-1} C_{21.0}}{C_{11}},$$

i.e.

$$U = U r_t^2 + U(1 - r_t^2), \tag{3·4}$$

where

$$r_t^2 = \frac{(\alpha' C_{22} a_t)^2}{(\alpha' C_{22} \alpha)(a_t' C_{22} a_t)} = \frac{\alpha' C_{21} C_{11}^{-1} C_{12} \alpha}{U(\alpha' C_{22} \alpha)} \tag{3·5}$$

is the square of the (total) correlation between the observed and the hypothetical discriminant functions. While no corresponding equation to (2·9) exists for Ω, we write similarly

$$\Omega = \Omega r_i^2 + \Omega(1 - r_i^2), \tag{3·6}$$

where

$$r_i^2 = \frac{(\alpha' C_{22.1} a_i)^2}{(\alpha' C_{22.1} \alpha)(a_i' C_{22.1} a_i)} \tag{3·7}$$

is the square of the "internal correlation" of the observed and hypothetical discriminant functions after elimination of S_1. Ωr_i^2 is equivalent to $U r_t^2$ in that it provides the same exact test of significance of the relation of $\alpha' S_2$ with S_1.

For testing the significance of the relation of the remainder of S_2 with S_1, neither (3·4) nor (3·6), however, provides the appropriate exact test, which we have already obtained in terms of the Λ criterion. This test is based on the relation between the reduced sample $S_{2.0}$ orthogonal to $S_0 \equiv \alpha' S_2$, and the vector $S_{1.0}$ *in the same space*, orthogonal to S_0. In terms of the formal regression analysis of S_1 on S_2, we should write

$$S_1 = B_{10} S_0 + B_{12.0} S_{2.0} + S_{1.2}, \tag{3·8}$$

and test the significance of $B_{12.0} S_{2.0}$ in comparison with $S_{1.2}$. The appropriate modification of the second term of (3·4) gives

$$\frac{U(1 - r_t^2) C_{11}}{C_{11.0}} = \frac{U(1 - r_t^2)}{1 - U r_t^2} = U(1 - r_i^2), \tag{3·9}$$

the last step following from equations (3·5) and (3·7); the corresponding criterion replacing Ω is

$$\Omega(1 - r_t^2). \tag{3·10}$$

Thus instead either of $U(1 - r_t^2)$ or $\Omega(1 - r_i^2)$ (the latter criterion being put forward by Fisher(4), p. 385), these two different criteria arising out of a formal additive analysis of U and Ω respectively, we have the criteria $U(1-r_i^2)$ and $\Omega(1-r_i^2)$, which are now equivalent and correspond to the Λ criterion. We have, for example,

$$\Omega(1-r_i^2) = a_i' C_{22.1} a_i (1-r_i^2)$$

in the form $(p-1)\,s_1^2/\{(n-p)\,s_2^2\}$ in relation to a z-test for the ratio of two variances s_1^2 and s_2^2 with $p-1$ and $n-p$ degrees of freedom respectively.

REFERENCES

(1) AITKEN, A. C. *Proc. Edinburgh Math. Soc.* 4 (1935), 106.
(2) BARTLETT, M. S. *Proc. Cambridge Phil. Soc.* 30 (1934), 327.
(3) BARTLETT, M. S. *Proc. Cambridge Phil. Soc.* 34 (1938), 33.
(4) FISHER, R. A. *Annals of Eugenics*, 8 (1938), 376.
(5) HOTELLING, H. *Biometrika*, 28 (1936), 321.
(6) LAWLEY, D. N. *Biometrika*, 30 (1938–9), 180 and 467.
(7) WILKS, S. S. *Biometrika*, 24 (1932), 471.

QUEENS' COLLEGE
CAMBRIDGE