The vector representation of a sample. By M. S. BARTLETT, B.A., Queens' College. (Communicated by Mr F. P. WHITE.)

[*Received* 8 May, *read* 28 May 1934.]

1. *Vector notation.* The geometrical or vectorial representation of a sample as a vector with n mutually perpendicular components corresponding to the n observations in the sample was introduced into statistics by Fisher (1), and led to the solution of many theoretical problems of statistical distributions. Subsequently Fisher (2) gave an alternative algebraic method applicable to finding distributions in connection with regression and the analysis of variance. The actual use of symbolic vector notation in deducing some of the properties of the sample vector—besides tending to stress the complementary character of these two different methods of proof, and the common principle underlying the analysis of any sample into its components—indicates also some points in connection with the assumption of the normal law in which it seems almost essential to retain the geometrical side of this vector representation. Moreover, it will be seen that the vector theory used here in reviewing briefly the analysis of a sample of one dependent variate can readily be extended to cover the case of correlated variates.

A sample of n observations of a variate x we may denote by

$$S = (x) = (x^{(1)}, \ldots, x^{(n)}).$$

For two variates x_1 and x_2 we have vectors S_1 and S_2. For the inner or scalar product $\Sigma x_1 x_2$ one common notation is $(S_1 S_2)$; but since we may regard S_1 as a matrix with one row, and S_2' the transposed matrix of S_2 with one column, it will be convenient here to write this product $S_1 S_2'$, the usual matrix multiplication being understood. In particular $S_1 S_1' = \Sigma x_1^2 = S_1^2$, say. The advantage of this latter notation is that when we consider, as an extension of S, the matrix of two vectors

$$S = \begin{pmatrix} S_1 \\ S_2 \end{pmatrix} = \begin{pmatrix} x_1^{(1)}, \ldots, x_1^{(n)} \\ x_2^{(1)}, \ldots, x_2^{(n)} \end{pmatrix}$$

for correlated variates x_1 and x_2, we have

$$|SS'| = S^2, \text{ say}, = \begin{vmatrix} S_1^2 & S_1 S_2' \\ S_1 S_2' & S_2^2 \end{vmatrix},$$

which is related to the generalised variance considered by Wilks (3); and similarly of course for any number of variates.

We may derive here an equation which will prove to be of value later (see § 8). For the case of one variate, we have the

https://doi.org/10.1017/S0305004100012512 Published online by Cambridge University Press

trivial relation that if $T = aS$, then $T^2 = a^2S^2$. If in general for any number of variates, we take any other linearly independent set of vectors in the "plane" (see footnote in § 2) of $S_1, S_2, \ldots$, given by

$$T = AS,$$

where A is the square matrix (a_{ij}) of coefficients, we have

$$\begin{aligned} |TT'| &= |AS(AS)'| \\ &= |ASS'A'| \\ &= |A|^2 |SS'|, \end{aligned}$$

or

$$T^2 = |A|^2 S^2. \tag{1}$$

2. *The analysis of a sample of one dependent variate.* To return to the simple vector S for one variate x, the most elementary analysis of S is to isolate the arithmetic mean $M = (\bar{x})$, and write

$$S = V + M,$$

where $V = (x - \bar{x})$. Squaring, we have

$$S^2 = (V + M)^2 = V^2 + M^2, \tag{2}$$

since

$$(VM) \equiv VM' = \Sigma \bar{x}(x - \bar{x}) = 0.$$

The fact that V and M are perpendicular denotes of course that they are *algebraically* independent, in the sense that any change which we insert in the mean M will not affect the variation V about the mean.

As another example of the analysis of S on more complicated but similar lines, we may consider the Latin square as used in agricultural field experiments (see (4), p. 246), where we have a classification in rows, columns and treatments. We write

$$S = R + C + T + E + M,$$

where $R = (\bar{x}_r - \bar{x})$ is the vector representing the differences of row means from the general mean, $C = (\bar{x}_c - \bar{x})$ similarly for columns, $T = (\bar{x}_t - \bar{x})$ for treatments, $M = (\bar{x})$ as before, and

$$E = (x - \bar{x}_r - \bar{x}_c - \bar{x}_t + 2\bar{x})$$

is the residual variation or error term. From the algebraic relations

$$RC' = RT' = \ldots = EM' = 0,$$

we have, analogously to (2),

$$S^2 = R^2 + C^2 + T^2 + E^2 + M^2. \tag{3}$$

For a composite sample made up of k independent samples, we can always write

$$S = \sum^{k} S_{(r)},$$

where $S_{(r)}$ is a vector with zero elements except for the group, n_r

https://doi.org/10.1017/S0305004100012512 Published online by Cambridge University Press

in number, corresponding to the observations of the rth sample, and naturally

$$S^2 = \sum^{k} S_{(r)}^2,$$

which we can always analyse further into

$$S^2 = \sum^{k} V_{(r)}^2 + \sum^{k} M_{(r)}^2,$$

corresponding to (2), or if

$$V = \sum^{k} V_{(r)}, \quad D = \sum^{k} M_{(r)} - M, \quad M = (\bar{x}),$$

then

$$S^2 = V^2 + D^2 + M^2. \tag{4}$$

The perfectly general case is where we regard a vector S as being related to a matrix Z of vectors $Z_1, \ldots, Z_p$, by a linear regression equation of the form

$$S = BZ + V, \tag{5}$$

where B is a single row matrix of coefficients b_i $(i = 1, \ldots, p)$. That is, we have $BZ = U$, say, as the component of S in the "plane" of $Z_1, \ldots, Z_p$, and V the residual component*. Since V is then perpendicular to this "plane", the coefficients b_i are identical with those that would be obtained by the method of least squares. The formal solution may be written down at once from (5), for we have

$$SZ' = BZZ' + VZ' = BZZ',$$

since $VZ' = 0$, and hence

$$B = (SZ')(ZZ')^{-1},$$

though the simplicity of this equation does not of course imply a corresponding simplicity in practice in the arithmetical labour of solution.

Finally writing $U = X + Y$, where X is the component in the restricted "plane" of vectors $Z_1, \ldots, Z_q$, $(q \leqslant p)$, which we wish to eliminate, and Y the remaining component, we have

$$S = X + Y + V, \tag{6}$$

where

$$S^2 = X^2 + Y^2 + V^2. \tag{7}$$

Corresponding to the isolation of any component contribution Y^2 to the total variation represented by S^2, there will always exist an equation of the form (7). In the straightforward theory of regression, Y^2 will represent the variation due to a group of variates after we have already eliminated the effect of others represented by $Z_1, \ldots, Z_q$ (this to include the elimination of the mean). Equation (7) can be regarded, however, as covering all other cases of "analysis of variance". To illustrate this, consider the problem of isolating

* It is convenient in this paper to use the term "plane" to denote the vector subspace determined by a fixed set of vectors. In equation (5) above, for example, the "plane" of Z is the space traversed by the vector BZ as B varies.

https://doi.org/10.1017/S0305004100012512 Published online by Cambridge University Press

the variation due to the difference in the observed regression coefficients of x on another variate y for k independent groups or samples (cf. (5), p. 168).

Take Z_r so that $M_{(r)} = \bar{x}_{(r)} Z_r$, $(r \leqslant k)$, where $M_{(r)}$ is the vector representing the mean of the rth sample. Corresponding to $S = V + M$ for the variate x, let $T = W + N$ refer to the variate y. Then take also $Z_{k+1} = W$; $Z_{k+2}, \ldots, Z_{2k}, = W_{(1)}, \ldots, W_{(k-1)}$; and note that $W_{(k)}$ is also a function of the Z_r's. We have

$$S = \sum^{k} M_{(r)} + \sum^{k} b_{(r)} W_{(r)} + E, \text{ say,}$$

where $b_{(r)}$ is the regression coefficient of x on y for the rth sample separately, and this can be written

$$S = \sum^{k} M_{(r)} + b W + \sum^{k} (b_{(r)} - b) W_{(r)} + E,$$

where b is the regression coefficient of x on y for the samples together (the sample means still, however, being eliminated), so that

$$b = \sum^{k} S_{(r)} V'_{(r)} / \sum^{k} V^2_{(r)}.$$

This is of the form required in (6), with

$$X = \sum^{k} M_{(r)} + b W, \quad Y = \sum^{k} (b_{(r)} - b) W_{(r)},$$

and

$$S^2 = X^2 + Y^2 + E^2, \tag{8}$$

where Y^2 is the quantity required.

3. *The vector condition for the normal law of error.* So far the properties proved above for the components of the vector S are algebraic or arithmetical properties—and illustrate, for example, in the case of the Latin square, that any actual differences in the row or column means will not affect either the treatment means or the error term, or in the case of regression, that any real linear relationship between S and a vector T will not affect the partial variation of S when T is given. Moreover, we have the important equations concerning mean values. For let the chance part of the vector S measured from its true mean be S_0, so that

$$S = C + S_0, \text{ say.}$$

The component U of S on any single vector Z is

$$(SZ'/Z^2)\, Z \quad \text{and} \quad U^2 = (SZ')^2/Z^2.$$

If the component of C is G, we have the expected value* of U^2,

$$E\{U^2\} = G^2 + E\{(S_0 Z)^2/Z^2\},$$

* The symbol E for expectation should not be confused with the vector and matrix notation.

https://doi.org/10.1017/S0305004100012512 Published online by Cambridge University Press

since $E\{S_0 Z'\} = 0$, and using the results

$$E\{x_0^{(r)} x_0^{(s)}\} = \begin{cases} 0, & (r \neq s) \\ \sigma^2, & (r = s) \end{cases}$$

we may easily prove algebraically the well-known result that

$$E\{U^2\} = G^2 + \sigma^2, \tag{9}$$

where σ^2 is the variance. Since the component of S in the "plane" of r linearly independent vectors is the sum of r perpendicular components, we obtain the formula for mean values corresponding to equations (6) and (7),

$$E\{S^2\} = (A^2 + q\sigma^2) + \{B^2 + (p-q)\sigma^2\} + \{(n-p)\sigma^2\}, \tag{10}$$

where $A + B$ is the total component of S in the "plane" of the vectors $Z_1, \ldots, Z_p$, other than the component due to chance. This result, especially for the last term, is a familiar one in the theory of least squares. Particular cases of it were considered by Irwin in connection with the "analysis of variance" (6).

To proceed further on the usual assumption of normality, we require the vector condition for the normal law, this being the very simple one that the chance component of the vector S is equally likely to have any direction in the n-dimensional vector space which we are considering. This condition, besides being necessary for the normal law, is easily seen to be sufficient; for, if the chance of a vector S is independent of its direction and depends only on its length, we may write

$$p(S) \propto f(S^2), \tag{11}$$

whence, writing $p(S) = \Pi\, p(x)$, taking logarithms, and differentiating for S^2 constant, we have

$$\Sigma \frac{\partial p(x)}{\partial x} \frac{dx}{p(x)} = 0$$

for

$$\Sigma x\, dx = 0,$$

whence

$$\frac{\partial p(x)}{\partial x} \Big/ p(x) + \lambda x = 0,$$

i.e.

$$p(x) \propto e^{-\frac{1}{2}\lambda x^2},$$

where λ is necessarily positive. Note also that it must be constant for each observation in the sample.

In general, for any number of correlated variates, if $p(S)$ depends only on the elements in the matrix SS', so that we may write symbolically

$$p(S) \propto f(SS'),$$

https://doi.org/10.1017/S0305004100012512 Published online by Cambridge University Press

then it may similarly be shown that these variates must all be normal.

It may be noted that equation (11) gives a simple proof of the property of the normal law that any linear function of n independent normal variates itself obeys the normal law, for we see that the random vector S is likely to bear exactly the same angular relation to any assigned line as it does to one of the original axes, and this line will correspond to a linear transformation of the original variates. Here we have assumed that our linear function is one of variates which may be regarded as observations in a sample, and hence have equal variances, but for the case of different variances we can always divide each variate by its standard deviation, at the same time incorporating the standard deviation in the coefficient; that is,

$$\Sigma ax = \Sigma (a\sigma)(x/\sigma).$$

We note further that only for the normal law will this property be true.

4. If we make the stipulation, therefore, that the part of the vector S depending on the chance variation of the observations is equally likely to have any direction, it follows (to repeat Fisher's original arguments) that it bears the same relationship to the component subspaces corresponding to the components of S as it does to the subspaces corresponding to different numbers of the original observations. Thus the component vectors are *absolutely* independent, in the probability sense, as well as *algebraically* independent, and the distribution of the component square X^2, for example, in equation (7), on the hypothesis that this square contains no real component A^2, will follow the distribution of $\chi^2\sigma^2$, with the number of degrees of freedom q corresponding to the number of dimensions in the vector space for this particular component.

The absolute independence of the vector components, on the assumption of the normal law, thus follows merely as a corollary of their algebraic independence. In fact, if for any reason in analysis of variance problems some of the simpler algebraic relations indicating perpendicularity are no longer available, any other independence properties which we can construct are naturally of no value unless they ensure the isolation of any component in which we are interested. We can always, however, use the general procedure of fitting regression constants by the theory of least squares, and obtain an equation of the form (7), where Y is the component which we wish to compare with the error term V, the component X having been eliminated. This has been clearly stated in connection with the analysis of agricultural field experiments by Yates (7), who considers also the use of approximate methods in certain cases where the exact process of fitting would be too laborious.

https://doi.org/10.1017/S0305004100012512 Published online by Cambridge University Press

5. To illustrate the interpretation of the equation

$$S = X + Y + V,$$

and the derivation of the exact test of significance of the component Y, in a problem where this exact test is readily obtained, consider an agricultural experiment where we may require to test the effect of treatments when not only heterogeneous variation due to different blocks or rows has to be eliminated, but also the regression of the variate in question on some other variate such as the yields of the plots at some preceding time. The component X is thus to include not only the general mean and the differences from this mean due to blocks or rows, but also the partial component of S on the vector representing this "independent variate". To find V we have to eliminate also differences from the general mean due to treatments. Consider for example the following data, used by Fisher ((4), p. 257) to illustrate the great increase in accuracy possible by allowing for these previous yields.

	Degrees of freedom	x^2	xy	y^2
Rows	3	745·0	837·0	1095·5
Columns	3	213·5	120·75	69·5
Error	9	567·5	654·25	875·0
Total	15	1526·0	1612·00	2040·0

In this table of sums of squares and products, y, the dependent variate, is an experimental tea-yield, and x the preliminary yield (thus S denotes here the sample for the variate y). Since there is no line for treatments, we may consider instead the testing of the component corresponding to columns. In his book Fisher does not give any test, but evidently such a test of significance will often be required. I found on discussing the point with him that he was aware of the test I give below, and that this is included in the draft for the fifth edition of his book, to be published during 1934.

It will be seen that the data are presented at first as a sample of two correlated variates (cf. (8); see also the footnote in § 7). Since in the sum of squares for y in the line for error all effects other than the regression on x have been eliminated, we obtain our adjusted error term V^2 by eliminating this effect also, using the figures in the line for error only. Similarly we find $Y^2 + V^2$ by eliminating the effect of x from the *totals* of columns and error,

https://doi.org/10.1017/S0305004100012512 Published online by Cambridge University Press

V^2 being then obtained by subtraction. We thus have the following analysis:

	Degrees of freedom	x^2	xy	y^2	b	Degrees of freedom	y_x^2
Columns	3	213·5	120·75	69·5		3	54·8
Error	9	567·5	654·25	875·0	1·1529	8	120·7
Total	12	781·0	775·00	944·5	·9923	11	175·5

It is apparent without any test being made that the mean square 18·27 to be used in testing the significance of the effect of columns is not significantly different from the corresponding mean square 15·09 for error.

6. One interesting theoretical point which the vector approach seems to make is the *necessity* of normality for the derived t and z distributions of "Student" and Fisher. For if we write

$$S = U + V,$$

where

$$S^2 = U^2 + V^2,$$

then the ratio U^2/S^2 (which determines the value of z) specifies the angle between S and its component U, and the probability distribution of this angle will hold only if the direction of S is completely random*. The corresponding algebraic approach, which shows that U^2 and V^2 are independent because they are made up of uncorrelated linear functions of the original observations, gives the distribution of U^2/V^2 or U^2/S^2, but does not give the necessary condition for this distribution to hold.

We at once obtain further the fact that S^2 and U^2/S^2 are independent, as mentioned elsewhere (5), since S^2 measures the length of S, and has no connection with the angle between S and U. Again this property will only be true when the length of the vector does not influence its direction, that is, for the normal law.

7. *Correlated variates*†. For two or more correlated variates, S was defined in § 1 as the matrix representing the set of vectors $S_1, S_2, \ldots$. In the geometrical arguments which have been used in connection

* Or alternatively if the "plane" in which U lies is random—see the theory of correlated variates.

† The term "correlated variates" implies here variates treated equally with each other, such, for example, as physical measurements in anthropometric and biological statistics.

Note that the statistical method known as "analysis of covariance" (see (8), and (4), p. 257) is often used to eliminate the effect of one or more variates on one

https://doi.org/10.1017/S0305004100012512 Published online by Cambridge University Press

with distributions in the theory of correlation (see (1), for example), the joint variation of these vectors was the relevant feature. It is of value also, however, to extend the general theory given above in vector form for one variate, and consider the analysis of the matrices S and SS' into components.

Corresponding to equation (5), we write

$$S = U + V, \tag{12}$$

where $U = BZ$ is the component of S in the "plane" of vectors $Z_1, \ldots, Z_l$, B being now a rectangular matrix of coefficients b_{ij} $(i = 1, \ldots, p, j = 1, \ldots, l)$ if l is the number of variates in S. We then have also the equation

$$SS' = UU' + VV', \tag{13}$$

where it must be remembered that this is a matrix equation, and does not imply that S^2 is equal to the sum of U^2 and V^2, where S^2 was defined as $|SS'|$. For example, in the case of two variates, it represents the three identities

$$\begin{cases} S_1^2 = U_1^2 + V_1^2, \\ S_1 S_2' = U_1 U_2' + V_1 V_2', \\ S_2^2 = U_2^2 + V_2^2. \end{cases}$$

Corresponding to the general equation (6), where we wish to eliminate the component in the "plane" of vectors $Z_1, \ldots, Z_q$, we have

$$S = X + Y + V, \tag{14}$$

where

$$SS' = XX' + YY' + VV'. \tag{15}$$

Corresponding to equation (10) for mean or expected values we have

$$E\{SS'\} = (AA' + q\Sigma) + \{BB' + (p-q)\Sigma\} + \{(n-p)\Sigma\}, \tag{16}$$

where Σ represents the matrix $(\rho_{ij}\sigma_i\sigma_j)$, and $A + B$ as before is the total non-chance component of S in the "plane" of $Z_1, \ldots, Z_p$.

The only additional result needed to establish this equation is one similar to (9),

$$E\{U_1 U_2'\} = G_1 G_2' + \rho_{12}\sigma_1\sigma_2 \tag{17}$$

for variates x_1 and x_2, say, where here U is the component of S on a simple vector Z (the proof being similar).

Again when we assume the normal law for the distribution of $x_1, x_2, \ldots$, the algebraic independence of X, Y and V ensures their absolute independence, so that the components XX', YY', and VV' in (15) are absolutely independent.

dependent variate, as illustrated in § 5. In that case, though we may make use of the algebraic identities available for the analysis of a sample of more than one variate, our analysis will be covered by the general equations for one variate, (6) and (7).

https://doi.org/10.1017/S0305004100012512 Published online by Cambridge University Press

8. Corresponding to the ratio $U^2/S^2 = \cos^2\theta$ in the case of one variate, where θ is the angle between S and its component V in the "plane" of $Z_1, \ldots, Z_p$, let us first consider the ratio $|UU'|/|SS'|$, defined as U^2/S^2. It should be noted that for any other set of l vectors in the "plane" of $S_1, S_2, \ldots$ given by $T = AS$, say, we shall have

$$T = AU + AV,$$

and from equation (1), since $(AU)^2/T^2 = U^2/S^2$, it follows that the ratio U^2/S^2 is invariant for any alternative set of l linearly independent vectors in the "plane" of S. Similarly V^2/S^2 is invariant.

Notice, however, that the "plane" of U has p dimensions, that of V usually the same number as the number l of variates in S. Suppose that $l > p$. Then we may write $U = BZ$, where for the moment Z is taken to be a set of vectors $Z_1, \ldots, Z_l$, and B thus a square matrix. Using equation (1) again, we have

$$U^2 = |B|^2 Z^2.$$

But since U is the component of S in the "plane" of $Z_1, \ldots, Z_p$, all coefficients b_{ij} in B are zero for $j > p$, whence $|B| = 0$, and $U^2 = 0$. Thus U^2 and U^2/S^2 are always zero if the number of correlated variates l exceeds p. If we require a measure of the "inclination" of the "plane" of S to the "plane" of U, it is evident therefore that U^2/S^2 is unsatisfactory, for there is no reason to restrict the number of correlated variates considered; moreover, even if $l < p$, this measure would not be a reciprocal one, for then the "inclination" of the "plane" of U to the "plane" of S would be zero. Since the "plane" of V in general has the same number of dimensions as the "plane" of S, the same criticism does not apply to the ratio V^2/S^2. We may thus use the quantity $1 - V^2/S^2$ as a measure of the "inclination" of the "plane" of S to the "plane" of U, this being 1 if $V = 0$, and 0 if $U = 0$.

This measure is moreover a reciprocal one. For let CS be the component of U in the "plane" of S. Then, since $S = U + V$,

$$U = CS + (1 - C)S - V,$$

where $$\{(1 - C)S - V\}(CS)' = 0,$$

that is, $$(1 - C)SS'C' = VV'C',$$

whence $$1 - C = (VV')(SS')^{-1}.$$

The reciprocal measure corresponding to V^2/S^2 can be written

$$|(U - CS)U'|/U^2 = |1 - C|$$
$$= V^2/S^2.$$

The measure $1 - V^2/S^2$ of the "inclination" of the "plane" of S to the "plane" of U is also, therefore, that of the "inclination" of the "plane" of U to the "plane" of S.

https://doi.org/10.1017/S0305004100012512 Published online by Cambridge University Press

If we are interested in the component U, and require a single criterion to gauge its magnitude, the quantity $1 - V^2/S^2$, or simply the ratio V^2/S^2, thus presents itself as a natural quantity to consider. More generally, when a component X is first eliminated, so that

$$S - X = Y + V,$$

the appropriate criterion for the component Y will be $V^2/(Y+V)^2$. It was mentioned that in general the "plane" of V has l dimensions. This will not hold if the number of degrees of freedom $n-p$ for V is less than l. Thus V^2 and $V^2/(Y+V)^2$ are zero if $l > n-p$, and we must always have $n \geqslant p+l$.

When we wish to make a *single* test of the significance of the component Y, we may thus use the ratio $V^2/(Y+V)^2$. It is not suggested that the above discussion is altogether sufficient to establish this ratio as the appropriate criterion to use, though in this vector method of approach no other seems available. Reference may also be made to a paper by Pearson and Wilks (9), where one problem they consider in detail is testing the differences in means for k samples of two variates. In this case,

$$S = M + D + V,$$

say (cf. equation (4) for one variate), and the ratio $V^2/(D+V)^2$ they show to be the λ likelihood criterion (raised to the $2/n$th power) that the means are all equal.

We may consider also the problem of testing whether the variates $x_1, \ldots, x_l$ depend on other variates represented by vectors $Z_2, \ldots$ (the vector Z_1 being used to eliminate the mean M). Now Wilks (3) has considered the general problem of testing whether groups of normal variates are correlated between groups, and gives the moments and distribution of the appropriate λ criterion. Taking the case of two groups, and regarding it more generally as a regression problem, so that we only assume the normality of the "dependent" group, we obtain the problem mentioned here, and the ratio $V^2/(Y+V)^2$, where the component X eliminated in this case is the set of means M, is then readily shown to be identical with the quantity Wilks considers.

It is evident that since the moments of the ratio

$$\Lambda = V^2/(Y+V)^2 \qquad (18)$$

are known in particular cases, they are obtained in the general case corresponding to equation (14) merely by adjusting the different numbers of degrees of freedom. Alternatively, they can always be obtained by factorising quantities like V^2, $(Y+V)^2$, etc. (thus S^2 for two variates is $S_1^2 S_{2.1}^2$, where $S_{2.1}$ is the component of S_2 perpendicular to S_1 and thus has one less degree of freedom).

The case of two variates is particularly simple, because the distribution of $\sqrt{\Lambda}$ is then always of a well-known form. Thus for

https://doi.org/10.1017/S0305004100012512 Published online by Cambridge University Press

testing the differences in means for k samples, Pearson and Wilks found that

$$p(\sqrt{\Lambda}) \propto (\sqrt{\Lambda})^{n-k-2}(1-\sqrt{\Lambda})^{k-2}\, d(\sqrt{\Lambda}). \tag{19}$$

Here, in equation (14), $q=1$, $p=k$.

For the regression problem considered above, from the moments given by Wilks, it is easily shown that

$$p(\sqrt{\Lambda}) \propto (\sqrt{\Lambda})^{n-r-3}(1-\sqrt{\Lambda})^{r-1}\, d(\sqrt{\Lambda}) \tag{20}$$

for r independent variates. Here $q=1$, $p=r+1$.

For the general case of equation (14), we have for two variates,

$$p(\sqrt{\Lambda}) \propto (\sqrt{\Lambda})^{n-p-2}(1-\sqrt{\Lambda})^{p-q-1}\, d(\sqrt{\Lambda}). \tag{21}$$

For the general case of l variates, the moments of Λ (obtained by factorisation) are given by

$$M_h(\Lambda) = E\{\Lambda^h\} = \prod_{i=0}^{l-1} \frac{\Gamma\{\frac{1}{2}(n-q-i)\}\,\Gamma\{\frac{1}{2}(n-p-i)+h\}}{\Gamma\{\frac{1}{2}(n-p-i)\}\,\Gamma\{\frac{1}{2}(n-q-i)+h\}}. \tag{22}$$

This expression simplifies in many instances. If $q=p-1$, $M_h(\Lambda)$ is analogous to the general $M_h(\Lambda)$ for $l=1$, and

$$p(\Lambda) \propto \Lambda^{\frac{1}{2}(n-l-q)-1}(1-\Lambda)^{\frac{1}{2}(l)-1}\, d\Lambda. \tag{23}$$

This corresponds to the test of a *single* factor, such, for example, as the set of means M ($q=0$), the difference in means between two samples ($q=1$), the regression of the variates on one independent variate ($q=1$), or the partial regression after s variates have been eliminated ($q=s+1$).

When $q=p-2$, we have $M_h(\Lambda)$ analogous to the general $M_h(\Lambda)$ for $l=2$, and $\sqrt{\Lambda}$ will have a distribution analogous to (21). This corresponds to the joint test of two factors.

It is of interest to observe that these simplifications are obvious on geometrical grounds. For the "inclination" of the "plane" of $S-X$ to the "plane" of the component Y was defined so that the measure was a reciprocal one. This "inclination" is random if *one* of the "planes" is random, and hence we have the theorem that the criterion $\Lambda(l, p-q)$ has the same distribution as the criterion $\Lambda(p-q, l)$, which is the criterion for one variate if $p=q+1$, for two variates if $p=q+2$.

9. To illustrate numerically the possible use of Λ in the case of two variates the following data may be considered*.

* These figures were used by Tippett ((**10**), p. 191) for illustrative purposes in connection with correlation. They were *selected* from a paper (**8**) giving the results of an experiment on the effect of fertilisers on the grain and straw yields of cereals (variates x_1 and x_2 above)—the yields of undressed plots being omitted. They may be assumed here to represent a complete experiment, being used in preference to the original data because the Λ test happens then to be less trivial. It will be understood that the figures are considered only in relation to this test.

https://doi.org/10.1017/S0305004100012512 Published online by Cambridge University Press

	Degrees of freedom	x_1^2	x_1x_2	x_2^2
Treatments	7	12,496·8	−6,786·6	32,985·0
Residual	49	136,972·6	58,549·0	71,496·1
(Total	56	149,469·4	51,762·4	104,481·1)
Blocks	7	86,045·8	56,073·6	75,841·5
Grand total	63	235,515·2	107,836·0	180,322·6

The last three columns denote the components of S_1^2, S_1S_2', S_2^2 respectively corresponding to the analysis $S = Y + V + X$ (except that the general mean M has been omitted). If a single test of the effect of treatments is required, we may find $V^2 = |VV'|$ from the "residual" row, $(V+Y)^2$ from the total of "treatments" and "residual", the "blocks" row being ignored altogether. We have

$$\sqrt{V^2} = 79780, \quad \sqrt{(V+Y)^2} = 113740, \quad \sqrt{\Lambda} = \cdot 7014.$$

From the form of the distribution of $\sqrt{\Lambda}$, we can clearly use the transformation

$$z = \tfrac{1}{2}\log_e\left\{\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}}\,\frac{2(n-p-1)}{2(p-q)}\right\}$$

if we wish to use Fisher's z test, n_1 and n_2 being taken to be $2(p-q)$ and $2(n-p-1)$. Here $n = 64$, $p = 15$, $q = 8$. Taking therefore

$$z = \tfrac{1}{2}\log_e\left(\frac{\cdot 2986}{\cdot 7014}\,\frac{96}{14}\right) = \cdot 536,$$

and noting that

$$z\,(1\,\%,\ n_1 = 14,\ n_2 = 96) = \cdot 406,$$

we conclude that the effect of treatments is clearly significant. We should therefore be justified in examining how each variate has contributed to this result. Since the mean square for x_1 in the treatment row is less than the mean square for error, and the significance of the treatment effect on x_2, though significant, is somewhat less so than the significance level obtained above (for x_2 only, $z = \cdot 586$ while $z\,(1\,\%,\ n_1 = 7,\ n_2 = 49) = \cdot 552$), it is of interest to note that the change in association between x_1 and x_2 in the treatment row as compared with the remaining figures has evidently contributed towards the significance level obtained.

In many cases it is obvious that the Λ test for two variates would be unnecessary—when for example the significance of the component in question for each variate separately is evident. But

https://doi.org/10.1017/S0305004100012512 Published online by Cambridge University Press

in doubtful cases it promises to be of some value, sometimes giving perhaps a significant result when neither variate taken separately does. It may therefore be stressed that it is available in all cases of the analysis of a sample of two variates for which the usual z test is employed to test the significance of a component for one variate. Note, however, that it may not be used to compare two different contributions to the error component V (this can be done in the case of one variate, though we notice that even here the test has strictly a different logical basis).

It will be seen that no consideration has been given to the practical use of the Λ test for three or more dependent variates (and $p - q > 2$), but in the absence of practical problems of this more complicated nature, this omission seems justified.

10. *Summary.* Vector notation is used in reviewing the analysis of a sample of one dependent variate, the general equation being of the form

$$S = X + Y + V,$$

where

$$S^2 = X^2 + Y^2 + V^2.$$

The vector condition for the normal law is considered, and its relation to the distribution of U^2/S^2 if $S = U + V$ or more generally of $Y^2/(V + Y)^2$ if $S - X = Y + V$.

This vector theory is extended to cover the analysis of a sample of any number of correlated variates, S being then the matrix of vectors $S_1, S_2, \ldots$. We have a corresponding analysis

$$S = X + Y + V,$$

where

$$SS' = XX' + YY' + VV';$$

and if a single test of the significance of Y is required, the criterion $1 - \Lambda$, where

$$\Lambda = V^2/(Y + V)^2,$$

a square S^2 being defined as $|SS'|$.

REFERENCES.

(1) R. A. Fisher, *Biometrika*, 10 (1915), 507–521.
(2) R. A. Fisher, *Metron*, 5, No. 3 (1925), 90–104.
(3) S. S. Wilks, *Biometrika*, 24 (1932), 471–494.
(4) R. A. Fisher, *Statistical Methods for Research Workers* (4th ed. 1932).
(5) M. S. Bartlett, *Proc. Camb. Phil. Soc.* 30 (1934), 164–169.
(6) J. O. Irwin, *Journ. Roy. Stat. Soc.* 94 (1931), 284–300.
(7) F. Yates, *Journ. Agric. Sci.* 23 (1933), 108–145.
(8) T. Eden and R. A. Fisher, *Journ. Agric. Sci.* 17 (1927), 548–562.
(9) E. S. Pearson and S. S. Wilks, *Biometrika*, 25 (1933), 353–378.
(10) L. H. C. Tippett, *The Methods of Statistics* (1931).

https://doi.org/10.1017/S0305004100012512 Published online by Cambridge University Press