Relations Between Two Sets of Variates*

Harold Hotelling
Columbia University

1. The Correlation of Vectors. The Most Predictable Criterion and the Tetrad Difference

Concepts of correlation and regression may be applied not only to ordinary one-dimensional variates but also to variates of two or more dimensions. Marksmen side by side firing simultaneous shots at targets, so that the deviations are in part due to independent individual errors and in part to common causes such as wind, provide a familiar introduction to the theory of correlation; but only the correlation of the horizontal components is ordinarily discussed, whereas the complex consisting of horizontal and vertical deviations may be even more interesting. The wind at two places may be compared, using both components of the velocity in each place. A fluctuating vector is thus matched at each moment with another fluctuating vector. The study of individual differences in mental and physical traits calls for a detailed study of the relations between sets of correlated variates. For example the scores on a number of mental tests may be compared with physical measurements on the same persons. The questions then arise of determining the number and nature of the independent relations of mind and body shown by these data to exist, and of extracting from the multiplicity of correlations in the system suitable characterizations of these independent relations. As another example, the inheritance of intelligence in rats might be studied by applying not one but s different mental tests to N mothers and to a daughter

* Presented before the American Mathematical Society and the Institute of Mathematical Statisticians at Ann Arbor, September 12, 1935.

of each. Then $\frac{s(s-1)}{2}$ correlation coefficients could be determined, taking each of the mother-daughter pairs as one of the N cases. From these it would be possible to obtain a clearer knowledge as to just what components of mental ability are inherited than could be obtained from any single test.

Much attention has been given to the effects of the crops of various agricultural commodities on their respective prices, with a view to obtaining demand curves. The standard errors associated with such attempts, when calculated, have usually been found quite excessive. One reason for this unfortunate outcome has been the large portion of the variance of each commodity price attributable to crops of other commodities. Thus the consumption of wheat may be related as much to the prices of potatoes, rye, and barley as to that of wheat. The like is true of supply functions. It therefore seems appropriate that studies of demand and supply should be made by groups rather than by single commodities*. This is all the more important in view of the discovery that demand and supply curves provide altogether inadequate foundation for the discussion of related commodities, the ignoring of the effects of which has led to the acceptance as part of classical theory of results which are wrong not only quantitatively but qualitatively. It is logically as well as empirically necessary to replace the classical one-commodity type of analysis, relating for example to the incidence of taxation, utility, and consumers' surplus, by a simultaneous treatment of a multiplicity of commodities†.

The relations between two sets of variates with which we shall be concerned are those that remain invariant under internal linear transformations of each set separately. Such invariants are not affected by rotations of axes in the study of wind or of hits on a target, or by replacing mental test scores by an equal number of independently weighted sums of them for comparison with physical measurements. If measurements such as height to shoulder and difference in height of shoulder and top of head are replaced by shoulder height and stature, the invariant relations with other sets of variates will not be affected. In economics there are important linear transformations corresponding for example to the mixing of different grades of wheat in varying proportions‡. Both prices and quantities are then transformed linearly.

In this case, besides the invariants to be discussed in this paper, there will

* The only published study known to the writer of groups of commodities for which standard errors were calculated is the paper of Henry Schultz, "Interrelations of Demand," in *Journal of Political Economy*, Vol. XLI. pp. 468–512, August, 1933. Some at least of the coefficients obtained are significant.

† Harold Hotelling, "Edgeworth's Taxation Paradox and the Nature of Demand and Supply Functions" in *Journal of Political Economy*, Vol. XL, pp. 577–616, October, 1932, and "Demand Functions with Limited Budgets" in *Econometrica*, Vol. III, pp. 66–78, January, 1935.

‡ Harold Hotelling, "Spaces of Statistics and their Metrization" in *Science*, Vol. LXVII, pp. 149–150, February 10, 1928.

be other resulting from the fact that the transformation of quantities is not independent of that of the prices, but is contragredient to it. (Cf. Section 16 below.)

Sets of variates, which may also be regarded as many-dimensional variates, or as vector possessed of frequency distributions, have been investigated from several different standpoints. The work of Gauss on least squares and that of Bravais, Galton, Pearson, Yule and others on multivariate distributions and multiple correlation are early examples. In "The Generalization of Student's Ratio*," the writer has given a method of testing in a manner invariant under linear transformations, and with full statistical efficiency, the significance of sets of means, of regression coefficients, and of differences of means or regression coefficients. A procedure generalizing the analysis of variance to vectors has been applied to the study of the internal structure of cell by means of Brownian movements, for which the vectors representing displacements in consecutive fifteen-second intervals were compared with their resultants to demonstrate the presence of invisible obstructions restricting the movement†. Finally, S.S. Wilks has published important work on relations among two or more sets of variates which are invariant under internal linear transformations‡. Denoting by A, B and D respectively the determinants of sample correlations within a set of s variates, within a set of t variates, and in the set consisting of all these $s + t$ variates, the distribution of the statistic,

$$z = \frac{D}{AB} \tag{1.1}$$

was determined exactly by Wilks under the hyothesis that the distribution is normal, with no population correlation between any variate in one set and any in the other. Wilks also found distributions of analogous functions of three or more sets, and of other related statistics.

The statistic (1.1) is invariant under internal linear transformations of either set, as will be proved in Section 4. Another example of such a statistic is provided by the maximum multiple correlation with either set of a linear function of the other set, which has been the subject of a brief study§. This problem of finding, not only a best predictor among the linear functions of one set, but at the same time the function of the other set which is predicts most accurately, will be solved in Section 3 in a more symmetrical manner. When the influence of these two linear functions is eliminated by partial

* *Annals of Mathematical Statistics*, Vol. II. pp. 360–378, August, 1931.

† L.G.M. Baas-Becking, Henriette van de Sande Bakhuyzen, and Harold Hotelling, "The Physical State of Protoplasm" in *Verhandelingen der Koninklijke Akademie van Wetenschappen te Amsterdam*, Second Section, Vol. v. (1928).

‡ "Certain Generalizations in the Analysis of Variance" in *Biometrika*, Vol. XXIV. pp. 471–494, November, 1932.

§ Harold Hotelling, "The Most Predictable Criterion" in *Journal of Educational Psychology*, Vol. XXVI. pp. 139–142, February, 1935.

correlation, the process may be repeated with the residuals. In this way we may obtain a sequence of pairs of variates, and of correlations between them, which in the aggregate will fully characterize the invariant relations between the sets, in so far as these can be represented by correlation coefficients. They will be called *canonical variates* and *canonical correlations*. Every invariant under general linear internal transformations, such for example as z, will be seen to be a function of the canonical correlations.

Observations of the values taken in N cases by the components of two vectors constitute two matrices, each of N columns. If each vector has s components, then each matrix has s rows. In this case we may consider the correlation coefficient between the C_s^N-rowed determinants in one matrix and the corresponding determinants in the other. Since a linear transformation of the variates in either set effects a linear transformation of the rows of the matrix of observations, which merely multiplies all these determinants by the same constant, it is evident that the correlation coefficient thus calculated is invariant in absolute value. We shall call it the *vector correlation* or *vector correlation coefficient*, and denote it by q. When $s = 2$, if we call the variates of one set x_1, x_2, and those of the other x_3, x_4, and r_{ij} the correlation of x_i with x_j, then it is easy to deduce with the help of the theorems stated in Section 2 below that

$$q = \frac{r_{13}r_{24} - r_{14}r_{23}}{\sqrt{(1 - r_{12}^2)(1 - r_{34}^2)}} \tag{1.2}$$

For larger values of s, q will have as its numerator the determinant of correlations of variates in one set with variates in the other, and as its denominator the geometric mean of the two determinants of internal correlations. A generalization of q for sets with unequal numbers of components will be given in Section 4.

Corresponding to the correlation coefficient r between two simple variates, T.L. Kelley has defined the *alienation coefficient* as $\sqrt{1 - r^2}$. The square of the correlation coefficient between x and y is the fraction of the variance of y attributable to x, while the square of the alienation coefficient is the fraction independent of x. If we adopt this apportionment of variance as a basis of generalization, we shall be consistent in calling $\sqrt{z}$ the *vector alienation coefficient*.

If there exists a linear function of one set equal to a linear function of the other—if for example x_1 is identically equal to x_3—the expression (1.2) for q reduces to a partial correlation coefficient. If one set consists of a single variate and the other of two or more, the vector correlation coincides with the multiple correlation. If each set contains only one variate, q is the simple correlation between the two. Thus the vector correlation coefficient provides a generalization of several familiar concepts.

The numerator of (1.2), known as the tetrad difference or tetrad, has been of much concern to psychologists. The vanishing in the population of all the tetrads among a set of tests is a necessary condition for the theory, propounded by Spearman, that scores on the tests are made up of a component common in varying degrees to all of them, and of independent components specific to each. The vanishing of some but not all of the tetrads is a condition for certain variants of the situation*. The sampling errors of the tetrad have therefore received much attention. In dealing with them it has been thought necessary to ignore at least three types of error:

(1) The standard error formulae used are only asymptotically valid for very large samples, with no means of determining how large a sample is necessary.
(2) The assumption is made implicitly that the distribution of the tetrad is normal, though this cannot possibly be the case, since the range is finit†.
(3) Since the standard error formulae involve unknown population values, these are in practice replaced by sample values. No limit is known for the errors committed in this way.

Now it is evident that to test whether the population value of the tetrad is zero—the only value of interest—is the same thing as to test the vanishing of any multiple of the tetrad by a finite non-vanishing quantity. Wishart‡ considered the tetrad of covariances, which is simply the product of the tetrad of correlations by the four standard deviations. For this function he found exact values of the mean and standard error, thus eliminating the first source of error mentioned above.

The exact distribution of q found in Section 8 below may be used to test the vanishing of the tetrad, eliminating the first and second sources of error. Unfortunately even this distribution involves a parameter of the population, one of the canonical correlations, which must usually be estimated from the sample, introducing again an error of the third type. However there may be cases in which this one parameter will be known from theory or from a larger sample.

Now it will be shown that q is the product of the canonical correlations. Hence at least one of these correlations is zero if the tetrad is. Thus still another test of the same hypothesis may be made in this way. Now we shall obtain for a canonical correlation vanishing in the population the extremely

* Truman L. Kelley, *Crossroads in the Mind of Man*, Stanford University Press, 1928. This book, in addition to relevant test data and discussion, contains references to the extensive literature, a standard error formula for the tetrad, and other mathematical developments.

† The first proof that the distribution of the tetrad approaches normality for large samples was given by J.L. Doob in an article, "The Limiting Distributions of Certain Statistics," in the *Annals of Mathematical Statistics*, Vol. VI. pp. 160–169 (September, 1935). The proof is applicable only if the population value of z is different from unity, i.e. if the sets x_1, x_2 and x_3, x_4 are not completely independent. If they are completely independent, the limiting distribution is of the form $\frac{1}{2}ce^{-c|t|}dt$, as Doob showed. What the distribution of the tetrad is for any finite number of cases no one knows.

‡ "Sampling Errors in the Theory of Two Factors" in *British Journal of Psychology*, Vol. XIX. pp. 180–187 (1928).

simple standard error formula $\frac{1}{\sqrt{n}}$, involving no unknown parameter. Thus this test evades errors of the third kind, but is subject to those of the first two, although the second is somewhat mitigated by an ultimate approach to normality, since the canonical correlations satisfy the criterion for approach to normality derived by Doob in the article cited. Further research is needed to find an exact test involving no unknown parameter. The question of whether this is possible raises a very fundamental problem in the theory of statistical inference. We shall, however, find exact distributions appropriate for testing a variety of hypotheses.

2. Theorems on Determinants and Matrices

We shall have frequent occasion to refer to the following well-known theorem, the proofs of which parallel those of the multiplication theorem for determinants, and which might advantageously be used in expounding many parts of the theory of statistics:

Given two arrays, each composed of m rows and n columns ($m \leqslant n$). The determinant formed by multiplying the rows of one array by those of the other equals the sum of the products of the m-rowed determinants in the first array by the corresponding m-rowed determinants in the second.

When the two arrays are identical, we have the corollary that the symmetrical determinant of the products of rows by rows of an array of m rows and n columns ($m \leqslant n$) equals the sum of the squares of the m-rowed determinants in the array, and is therefore not negative.

3. Canonical Variates and Canonical Correlations. Applications to Algebra and Geometry

If x_1, x_2, ... are variates having zero expectations and finite covariances, we denote these covariances by

$$\sigma_{\alpha\beta} = Ex_{\alpha}x_{\beta},$$

where E stands for the mathematical expectation of the quantity following. If new variates $x'_1, x'_2, \ldots$ are introduced as linear functions of the old, such that

$$x'_{\gamma}\sum_{\alpha} c_{\gamma\alpha}x_{a},$$

then the covariances of the new variates are expressed in terms of those of the old by the equations

$$\sigma'_{\gamma\delta} = \sum_{\alpha\beta}\sum c_{\gamma\alpha} c_{\delta\beta} \sigma_{\alpha\beta} \tag{3.1}$$

obtained by substituting the equations above directly in the definition

$$\sigma'_{\gamma\delta} = Ex'_{\gamma} x'_{\delta},$$

and taking the expectation term by term.

Now (3.1) gives also the formula for the transformation of the coefficients of a quadratic form $\sum\sum \sigma_{\alpha\beta} x_\alpha x_\beta$ when the variables are subjected to a linear transformation. Hence the problem of standardizing the covariances among a set of variates by linear transformations is algebraically equivalent to the canonical reduction of a quadratic form. The transformation of the quadratic form into a sum of squares corresponds to replacing a set of variates by uncorrelated components. It is to be observed that the fundamental nature of covariances implies that $\sum\sum \sigma_{\alpha\beta} x_\alpha x_\beta$ is a positive definite quadratic form, and that only real transformations are relevant to statistical problems.

Considering two sets of variates $x_1, \ldots, x_s$ and $x_{s+1}, \ldots, x_{s+t}$, we shall denote the covariances, in the sense of expectations of products, by $\sigma_{\alpha\beta}$, $\sigma_{\alpha i}$, and σ_{ij}, using Greek subscripts for the indices $1, 2, \ldots, s$ and Latin subscripts for $s + 1, \ldots, s + t$. Determination of invariant relations between the two sets by means of the correlations or covariances among the $s + t$ variates is associated with the algebraic problem, which appears to be new, of determining the invariants of the system consisting of two positive definite quadratic forms

$$\sum_{\alpha\beta}\sum \sigma_{\alpha\beta} x_\alpha x_\beta, \qquad \sum_{ij}\sum \sigma_{ij} x_i x_j,$$

in two separate sets of variables, and of a bilinear form

$$\sum_{\alpha i}\sum \sigma_{\alpha i} x_\alpha x_i$$

in both sets, under real linear non-singular transformations of the two sets separately.

Sample covariances are also transformed by the formula (3.1). The ensuing analysis might therefore equally well be carried out for a sample instead of for the population. Correlations might be used instead of covariances, either for the sample or for the population, by introducing appropriate factors, or by assuming the standard deviations to be unity.

We shall assume that there is no fixed linear relation among the variates, so that the determinant of their covariances or correlations is not zero. This implies that there is no fixed linear relation among any subset of them; consequently every *principal* minor of the determinant of $s + t$ rows is different from zero.

If we consider a function u of the variates in the first set and a function v of those in the second, such that

$$u = \sum_\alpha a_\alpha x_\alpha, \qquad v = \sum_i b_i x_i,$$

the conditions

$$\sum\sum \sigma_{\alpha\beta} a_\alpha a_\beta = 1, \qquad \sum\sum \sigma_{ij} b_i b_j = 1 \tag{3.2}$$

are equivalent to requiring the standard deviations of u and v to be unity. The correlation of u with v is then

$$R = \sum_{\alpha i}\sum \sigma_{\alpha i} a_\alpha b_i. \tag{3.3}$$

If u and v are chosen so that this correlation is a maximum, the coefficients a_α and b_i will satisfy the equations obtained by differentiating

$$\sum\sum \sigma_{\alpha i} a_\alpha b_i - \tfrac{1}{2}\lambda \sum\sum \sigma_{\alpha\beta} a_\alpha a_\beta - \tfrac{1}{2}\mu \sum\sum \sigma_{ij} b_i b_j,$$

namely

$$\sum_i \sigma_{\alpha i} b_i - \lambda \sum_\beta \sigma_{\alpha\beta} a_\beta = 0 \tag{3.4}$$

$$\sum_\alpha \sigma_{\alpha i} a_\alpha - \mu \sum_j \sigma_{ij} b_j = 0. \tag{3.5}$$

Here λ and μ are Lagrange multipliers. Their interpretation will be evident upon multiplying (3.4) by a_α and summing with respect to α, then multiplying (3.5) by b_i and summing with respect to i. With (3.2) and (3.3), this process gives

$$\lambda = \mu = R.$$

The $s + t$ homogeneous linear equations (3.4) and (3.5) in the $s + t$ unknowns a_α and b_i will determine variates u and v making R a maximum, a minimum, or otherwise stationary, if their determinant vanishes. Since $\lambda = \mu$, this condition is

$$\begin{vmatrix} -\lambda\sigma_{11} \cdots -\lambda\sigma_{1s} & \sigma_{1,s+1} \cdots\cdots \sigma_{1,s+t} \\ \cdots\cdots\cdots & \cdots\cdots\cdots \\ -\lambda\sigma_{s1} \cdots -\lambda\sigma_{ss} & \sigma_{s,s+1} \cdots\cdots \sigma_{s,s+t} \\ \sigma_{s+1,1} \cdots \sigma_{s+1,s} & -\lambda\sigma_{s+1,s+1} \cdots -\lambda\sigma_{s+1,s+t} \\ \cdots\cdots\cdots & \cdots\cdots\cdots \\ \sigma_{s+t,1} \cdots \sigma_{s+t,s} & -\lambda\sigma_{s+t,s+1} \cdots -\lambda\sigma_{s+t,s+t} \end{vmatrix} = 0. \tag{3.6}$$

This symmetrical determinant is the discriminant of a quadratic form $\phi - \lambda\psi$, where

$$\phi = 2\sum_{\alpha i}\sum \sigma_{\alpha i} z_\alpha z_i, \qquad \psi = \sum_{\alpha\beta}\sum \sigma_{\alpha\beta} z_\alpha z_\beta + \sum_{ij}\sum \sigma_{ij} z_i z_j.$$

Here ψ is positive definite because it is the sum of two positive definite quadratic forms. Consequently* all the roots of (3.6) are real. Moreover the elementary divisors are all of the first degree†. This means that the matrix of

* Maxime Bôcher, *Introduction to Higher Algebra*, New York, 1931, p. 170, Theorem 1.

† Bôcher, p. 305, Theorem 4; p. 267, Theorem 2; p. 271, Definition 3.

the determinant in (3.6) is reducible, by transformations which do not affect either its rank or its linear factors, to a matrix having zeros everywhere except in the principal diagonal, while the elements in this diagonal are polynomials

$$E_1(\lambda), \qquad E_2(\lambda), \ldots, E_{s+t}(\lambda),$$

none of which contains any linear factor of the form $\lambda - \rho$ raised to a degree higher than the first. Therefore, if a simple root of (3.6) is substituted for λ, the rank is $s + t - 1$; but substitution of a root of multiplicity m for λ makes the rank $s + t - m$. Consequently if a simple root is substituted for λ and μ in (3.4) and (3.5) these equations will determine values of $a_1, a_2, \ldots, a_s, b_{s+1}, \ldots, b_{s+t}$, uniquely except for constant factors whose absolute values are determinable from (3.2). Not all these quantities are zero; from this fact, and the form of (3.4) and (3.5), it is evident that at least one a_α and at least one b_i differ from zero, provided the value put for λ is not zero. The variates u and v will then be fully determinate except that they may be replaced by the pair $-u$, $-v$. But for a root of multiplicity m there will be m linearly independent solutions instead of one in a complete set of solutions. From these may be obtained m different pairs of variates u and v.

The coefficient of the highest power of λ in (3.6) is the product of two principal minors, both of which differ from zero because the variates have been assumed algebraically independent. The equation is therefore of degree $s + t$. We assume as a mere matter of notation, if $s \neq t$, that $s < t$. Then of the $s + t$ roots at least $t - s$ vanish; for the coefficients of λ^{t-s-1} and lower powers of λ are sums of principal minors of $2s + 1$ or more rows, in which λ is replaced by zero, and every such minor vanishes, as can be seen by a Laplace expansion. Also, the sign of λ may be changed in (3.6) without changing the equation, for this may be accomplished by multiplying each of the first s rows and last t columns by -1. Therefore the negative of every root is also a root. The $s + t - (t - s) = 2s$ roots that do not necessarily vanish consist therefore of s positive or zero roots $\rho_1, \rho_2, \ldots, \rho_s$, and of the negatives of these roots. These s roots which are positive or zero we shall call the *canonical correlations* between the sets of variates; the corresponding linear functions u, v whose coefficients satisfy (3.2), (3.4) and (3.5) we call *canonical variates**. It is clear that every canonical correlation is the correlation coefficient between a pair of canonical variates. Hence no canonical correlation can exceed unity. The greatest canonical correlation is the maximum multiple correlation with either set of a disposable linear function of the other set. If u, v are canonical variates corresponding to ρ_γ, then the pair u, $-v$ or v, $-u$ is associated with the root $-\rho_\gamma$.

If a pair of canonical variates corresponding to a root ρ_γ is

$$u_\gamma = \sum_\alpha a_{\alpha\gamma} x_\alpha, \qquad v_\gamma = \sum_i b_{i\gamma} x_i, \tag{3.7}$$

* The word "canonical" is used in the algebraic theory of invariants with a meaning consistent with that of this paper.

the coefficients must satisfy (3.4) and (3.5), so that

$$\sum_i \sigma_{\alpha i} b_{i\gamma} = \rho_\gamma \sum_\beta \sigma_{\alpha\beta} a_{\beta\gamma}, \tag{3.8}$$

$$\sum_\alpha \sigma_{\alpha i} a_{\alpha\gamma} = \rho_\gamma \sum_j \sigma_{ij} b_{j\gamma}. \tag{3.9}$$

Also let

$$u_\delta = \sum_\beta a_{\beta\delta} x_\beta, \qquad v_\delta = \sum_i b_{i\delta} x_i \tag{3.10}$$

be canonical variates associated with a canonical correlation ρ_δ. Among the four variates (3.7) and (3.10) there are six correlations. Apart from ρ_γ and ρ_δ these are obviously

$$\left.\begin{array}{ll} Eu_\gamma u_\delta = \sum\sum \sigma_{\alpha\beta} a_{\beta\gamma} a_{\alpha\delta}, & Eu_\gamma v_\delta = \sum\sum \sigma_{\alpha i} a_{\alpha\gamma} b_{i\delta} \\ Ev_\gamma u_\delta = \sum\sum \sigma_{\alpha i} b_{i\gamma} a_{\alpha\delta}, & Ev_\gamma v_\delta = \sum\sum \sigma_{ij} b_{i\gamma} b_{j\delta} \end{array}\right\}. \tag{3.11}$$

We shall prove that the last four are all zero. Multiply (3.8) by $a_{\alpha\delta}$ and sum with respect to α. The result, with the help of (3.11), may be written

$$Ev_\gamma u_\delta = \rho_\gamma E u_\gamma u_\delta. \tag{3.12}$$

Multiplying (3.9) by $b_{i\delta}$ and summing with respect to i, we get

$$Eu_\gamma v_\delta = \rho_\gamma E v_\gamma u_\delta. \tag{3.13}$$

Interchanging γ and δ in this and then using (3.12), we obtain

$$\rho_\gamma E u_\gamma u_\delta = \rho_\delta E v_\gamma v_\delta. \tag{3.14}$$

Again interchanging γ and δ, we have

$$\rho_\delta E u_\gamma u_\delta = \rho_\gamma E v_\gamma v_\delta.$$

If $\rho_\gamma^2 \neq \rho_\delta^2$, the last two equations show that $Eu_\gamma u_\delta = Ev_\gamma v_\delta = 0$. Hence, by (3.12) and (3.13), $Ev_\gamma u_\delta$ and $Eu_\gamma v_\delta$ vanish. Thus all the correlations among canonical variates are zero except those between the canonical variates associated with the same canonical correlation.

If ρ_α is a root of multiplicity m, it is possible by well-known processes to obtain m solutions of the linear equations such that, if

$$a_{1\gamma}, \ldots, a_{s\gamma}, \quad b_{s+1,\gamma}, \ldots, b_{s+t,\gamma},$$

$$a_{1\delta}, \ldots, a_{s\delta}, \quad b_{s+1,\delta}, \ldots, b_{s+t,\delta},$$

are any two of these solutions, they will satisfy the orthogonality condition

$$\sum_\alpha a_{\alpha\gamma} a_{\alpha\delta} + \sum_i b_{i\gamma} b_{i\delta} = 0. \tag{3.15}$$

There is no loss of generality in supposing that each of the original variates was uncorrelated with the others in the same set and had unit variance. In this case (3.15) is equivalent to

$$Eu_\gamma u_\delta + Ev_\gamma v_\delta = 0,$$

where u_γ, v_γ, u_δ are given by (3.7) and (3.10). For this case of equal roots we have also from (3.14),

$$\rho_\alpha(Eu_\gamma u_\delta - Ev_\gamma v_\delta) = 0.$$

If $\rho_\alpha \neq 0$, the last two equations show that $Eu_\gamma u_\delta = Ev_\gamma v_\delta = 0$, and then from (3.12) and (3.13) we have that $Ev_\gamma u_\delta = Eu_\gamma v_\delta = 0$. These correlations also vanish if $\rho_\alpha = 0$, for then the right-hand members of (3.8) and (3.9) vanish, leaving two distinct sets of equations in disjunct sets of unknowns. The solutions may therefore be chosen so that the two sums in (3.15) vanish separately.

A double zero root determines uniquely, if $s = t$, a pair of canonical variates. If $s < t$, such a root determines a canonical variate for the less numerous set, and leaves $t - s$ degrees of freedom for the choice of the other.

The reduction of our sets of variates to canonical form may be completed by the choice of new variates $v_{s+1}, v_{s+2}, \ldots, v_t$ as linear functions of the second and more numerous set (unless the numbers in the two sets are equal), uncorrelated with each other and with the canonical variates v_γ previously determined, and having unit variance. This may be done in infinitely many ways, as is well known. These variates will also be uncorrelated with the canonical variates u_γ. Indeed, if

$$v_k = \sum b_{jk} x_j$$

is one of them, its correlation with u_γ is, by (3.7) and (3.9),

$$Eu_\gamma v_k = \sum\sum \sigma_{\alpha j} a_\gamma b_{jk} = \rho_\gamma \sum\sum \sigma_{ij} b_{i\gamma} b_{jk} = \rho_\gamma Ev_\gamma v_k,$$

which vanishes because v_k was defined to be uncorrelated with v_γ.

The normal form of two sets of variates under internal linear transformations is thus found to consist of linear functions $u_1, u_2, \ldots, u_s$ of one set, and $v_1, v_2, \ldots, v_t$ of the other, such that all the correlations among these linear functions are zero, except that the correlation of u_γ with v_γ is a positive number ρ_γ ($\gamma = 1, 2, \ldots, s$). Therefore *the only invariants of the system under internal linear transformations are $\rho_1, \rho_2, \ldots, \rho_s$, and functions of these quantities.*

The solution of the algebraic problem mentioned at the beginning of this section, by steps exactly parallel to those just taken with the statistical problem, is the following:

The positive definite quadratic forms $\sum\sum \sigma_{\alpha\beta} x_\alpha x_\beta$, and $\sum\sum \sigma_{ij} x_i x_j$, and the bilinear form $\sum\sum \sigma_{\alpha i} x_\alpha x_i$ with real coefficients, where the Latin subscripts are summed from 1 to s and the Greek subscripts from $s + 1$ to $s + t$, and $s \leqslant t$, may be reduced by a real linear transformation of $x_1, \ldots, x_s$ and a real linear transformation of $x_{s+1}, \ldots, x_{s+t}$ simultaneously to the respective forms $x_1^2 + \cdots + x_s^2$, $x_{s+1}^2 + \cdots + x_{s+t}^2$, and $\rho_1 x_1 x_{s+1} + \rho_2 x_2 x_{s+2} + \cdots + \rho_s x_s x_{2s}$. A fundamental system of invariants under such transformations consists of $\rho_1, \ldots, \rho_s$.

This algebraic theorem holds also if the quadratic forms are not restricted to be positive definite, provided (3.6) has no multiple roots and the forms are non-singular.

The normalization process we have defined may also be carried out for a sample, yielding canonical correlations $r_1, r_2, \ldots, r_s$, which may be regarded as estimates of $\rho_1, \rho_2, \ldots, \rho_s$, and associated canonical variates. With sampling problems raised in this way we shall largely be concerned in the remainder of this paper.

A further application is to geometry. In a space of N dimensions a sample of N values of a variate may be represented by a point whose coordinates are the observed values. The sample correlation between two variates is the cosine of the angle between lines drawn from the origin to the representing points, with the proviso, since deviations from means are used in the expression for a correlation, that the sum of all the coordinates of each point be zero. A sample of $s + t$ variates determines a flat space of s and one of t dimensions, intersecting at the origin, and containing the points representing the two sets of variates. In typical cases these two flat spaces do not intersect except at this one point. A complete set of metrical invariants of a pair of flat spaces is esaily seen from the foregoing analysis to consist of s angles whose cosines are $r_1, \ldots, r_s$. Indeed, like all correlations, they are invariant under rotations of the N-space about the origin, and they do not depend on the particular points used to define the two flat spaces. Each of these invariants is the angle between a line in one flat space and a line in the other. One of the invariants is the *minimum* angle of this kind, and the others are in a sense stationary. The condition that the two flat spaces intersect in a line is that one of the invariant quantities $r_1, \ldots, r_s$ be unity. They intersect in a plane if two of these quantities equal unity. For two planes through a point in space of four or more dimensions, there will be two invariants r_1, r_2, of which one is the cosine of the minimum angle. If $r_1 = r_2$, the planes are *isocline.* Every line in each plane then makes the minimum angle with some line in the other. If $r_1 = r_2 = 0$, the planes are completely perpendicular; every line in one plane is then perpendicular to every line in the other. If one of these invariants is zero and the other is not, the planes are semi-perpendicular; every line in each plane is perpendicular to a certain line in the other.

The determinant of the correlations among canonical variates is

$$\Delta = \begin{vmatrix} 1 & 0 & \ldots & 0 & \rho_1 & 0 & \ldots & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 & 0 & \rho_2 & \ldots & 0 & \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \ldots & 1 & 0 & 0 & \ldots & \rho_s & \ldots & 0 \\ \rho_1 & 0 & \ldots & 0 & 1 & 0 & \ldots & 0 & \ldots & 0 \\ 0 & \rho_2 & \cdots & 0 & 0 & 1 & \ldots & 0 & \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \ldots & \rho_s & 0 & 0 & \ldots & 1 & \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & \ldots & 1 \end{vmatrix}$$

$$= (1 - \rho_1^2)(1 - \rho_2^2)\ldots(1 - \rho_s^2). \tag{3.16}$$

The rank of the matrix

$$\left\| \begin{matrix} \rho_{1,s+1} \cdots\cdots\cdots \rho_{1,s+t} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \rho_{s,s+1} \cdots\cdots\cdots \rho_{s,s+t} \end{matrix} \right\|.$$

of correlations between the two sets is invariant under non-singular linear transformations of either set. Transformation to canonical variate reduces this matrix to

$$\left\| \begin{matrix} \rho_1 & 0 & \dots & 0 & \cdots\cdots\cdots & 0 \\ 0 & \rho_2 & \dots & 0 & \cdots\cdots\cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots\cdots\cdots & \cdots \\ 0 & 0 & \dots & \rho_s & \cdots\cdots\cdots & 0 \end{matrix} \right\|.$$

The rank is therefore the number of canonical correlations that do not vanish. This is the number of independent components common to the two sets. In the parlance of mental testing, the number of "common factors" of two sets of tests (e.g. mental and physical, or mathematical and linguistic tests) is the number of non-vanishing canonical correlations.

4. Vector Correlation and Alienation Coefficients

In terms of the covariances among the variates in the two sets $x_1, \dots, x_s$ and $x_{s+1}, \dots, x_{s+t}$, we define the following determinants, maintaining the convention that Greek subscripts take values from 1 to s, and Latin subscripts take values from $s + 1$ to $s + t$. It will be assumed throughout that $s \leqslant t$. A is the determinant of the covariances among the variates in the first set, arranged in order: that is, the element in the αth row and βth column of A is $\sigma_{\alpha\beta}$. B is the determinant of the covariances among variates in the second set, likewise ordered. D is the determinant of $s + t$ rows containing in order all the covariances among all the variates of both sets. C is obtained from D by replacing the covariances among the variates of the first set, including their variances, by zeros. Symbolically,

$$A = |\sigma_{\alpha\beta}|, \quad B = |\sigma_{ij}|, \quad C = \left| \begin{array}{c|c} 0 & \sigma_{i\alpha} \\ \hline \sigma_{\alpha i} & \sigma_{ij} \end{array} \right|, \quad D = \left| \begin{array}{c|c} \sigma_{\alpha\beta} & \sigma_{i\alpha} \\ \hline \sigma_{\alpha i} & \sigma_{ij} \end{array} \right|.$$

Suppose now that new variates $x'_1, \dots, x'_s$ are defined in terms of the old variates in the first set by the s equations

$$x'_\gamma = \sum c_{\gamma\alpha} x_\alpha.$$

The new covariances are then expressed in terms of the old by (3.1). The determinant of these new covariances, which we shall denote by A', may by

(3.1) and the multiplication theorem of determinants be expressed as the product of three determinants, of which two equal the determinant $c = |c_{\gamma\alpha}|$ of the coefficients of the transformation, while the third is A. If the variates of the second set are subjected to a transformation of determinant d, the determinants of covariances among the new variates analogous to those defined above are readily seen in this way to equal

$$A' = c^2 A, \quad B' = d^2 B, \quad C' = c^2 d^2 C, \quad D' = c^2 d^2 D. \tag{4.1}$$

Thus A, B, C, D are *relative invariants* under internal transformations of the two sets of variates.

The ratios

$$Q^2 = \frac{(-1)^s C}{AB} \quad \text{and} \quad Z = \frac{D}{AB} \tag{4.2}$$

we shall call respectively the squares of the *vector correlation coefficient* or *vector correlation*, and of the *vector alienation coefficient*. It is evident that both are absolute invariants under internal transformations of the two sets, since their values computed from transformed variates have numerators and denominators multiplied by the same factor $c^2 d^2$, in accordance with (4.1).

The notation just used is appropriate to a population, but the same definitions and reasoning may be applied to a sample. We denote by q^2 and z the same functions of the sample covariances that Q^2 and Z, respectively, have been defined to be of the population covariances.

A particularly simple linear transformation consists of dividing each variate by its standard deviation. The covariances among the new variates are then the same as their correlations, which are also the correlations among the old variates. Hence, in the definitions of the vector correlation and alienation coefficients, the covariances may be replaced by the correlations. For example, if $s = t = 2$, the squared vector correlation in a sample may be written

$$q^2 = \frac{\begin{vmatrix} 0 & 0 & r_{13} & r_{14} \\ 0 & 0 & r_{23} & r_{24} \\ r_{31} & r_{32} & 1 & r_{34} \\ r_{41} & r_{42} & r_{43} & 1 \end{vmatrix}}{\begin{vmatrix} 1 & r_{12} \\ r_{12} & 1 \end{vmatrix} \cdot \begin{vmatrix} 1 & r_{34} \\ r_{34} & 1 \end{vmatrix}} = \frac{(r_{13}r_{24} - r_{14}r_{23})^2}{(1 - r_{12}^2)(1 - r_{34}^2)}. \tag{4.3}$$

The vector correlation coefficient will always be taken as the positive square root of q^2 or of Q^2 (which are seen below to be positive) when $s < t$, and usually also when $s = t$. However, if in accordance with (4.3) we write

$$q = \frac{r_{13}r_{24} - r_{14}r_{23}}{\sqrt{(1 - r_{12}^2)(1 - r_{34}^2)}}, \tag{4.4}$$

it is evident that q may be positive for some samples of a particular set of variates, and negative for other samples. It may sometimes be advantageous, as in testing whether two samples arose from the same population, to retain the sign of q for each sample, since this provides evidence in addition to that given by the absolute value of q. But unless otherwise stated we shall always regard q as the positive root of q^2. Likewise, Q, $\sqrt{z}$ and $\sqrt{Z}$ will denote the positive roots unless otherwise specifically indicated in each case. A transformation of either set will reverse the sign of the algebraic expression (4.4) if the determinant of the transformation is negative. This will be true of a simple interchange of two variates; for example, $x'_1 = x_2$, $x'_2 = x_1$ has the determinant -1. On the other hand, the sign is conserved if the determinant of the transformation is positive. Such considerations apply whenever $s = t$.

Since the vector correlation and alienation coefficients are invariants, they may be computed on the assumption that the variates are canonical. In this case $A = B = 1$, and D is given by (3.16). To obtain C we replace the first s 1's in the principal diagonal of (3.16) by 0's. It then follows that

$$C = (-1)^s \rho_1^2 \rho_2^2 \dots \rho_s^2.$$

This confirms that the value of Q^2 given in (4.2) is positive. In this way the vector correlation and alienation coefficients are expressible in terms of the canonical correlations by the equations

$$Q = \pm \rho_1 \rho_2 \dots \rho_s, \qquad Z = (1 - \rho_1^2)(1 - \rho_2^2)\dots(1 - \rho_s^2), \tag{4.5}$$

$$q = \pm r_1 r_2 \dots r_s, \qquad z = (1 - r_1^2)(1 - r_2^2)\dots(1 - r_s^2). \tag{4.6}$$

From these results it is obvious that both the vector correlation and vector alienation coefficients are confined to values not exceeding unity. Also Z and z are necessarily positive, except that they vanish if, and only if, all the variates in one set are linear functions of those in the other.

Since the denominator of (4.4) is obviously less than unity, and since we have just shown that $q \leqslant 1$, the tetrad must be still less. This simple proof that the tetrad is between -1 and $+1$ shows the falsity of the idea that the range of the tetrad is from -2 to $+2$, which has gained some currency. An equivalent proof in vector notation was communicated to the writer by E.B. Wilson.

The only case in which Z can attain its maximum value unity is that in which all the canonical correlations vanish. In this case no variate in either set is correlated with any variate in the other, so that the two sets are completely independent, at least if the distribution is normal. Moreover, $Q = 0$. On the other hand, the only case in which Q can be unity is that in which all the canonical correlations are unity. In this event, $Z = 0$; also, the variates in the first set are linear functions of those in the second. Thus either z, $1 - q$, or $1 - q^2$ might be used as an index of independence, while we might use q, q^2 or $1 - z$ as a measure of relationship between the two sets. The work of Wilks alluded to in Section 1 provides an exact distribution of z on the hypothesis of complete independence, a distribution which may thus be used to test this hypothesis.

If we regard the elements of A, B and C as sample covariances, we have in case $s = t$ a simple interpretation of q. Consider the two matrices of observations on the two sets of variates in N individuals, in which each row corresponds to a variate and each column to an individual observed. From Section 2 it is evident that the square of the sum of the products of corresponding s-rowed determinants in the two matrices is $(-1)^s N^{2s} C$; also that the sums of squares of the s-rowed determinants in the two matrices are $N^s A$ and $N^s B$. Therefore q is simply the product-moment correlation coefficient between corresponding s-rowed determinants.

The generalized variance of a set of variates may be defined as the determinant of their ordered covariances, such as A or B. Let $\xi_1, \xi_2, \ldots, \xi_s$ be estimates respectively of $x_1, x_2, \ldots, x_s$ obtained from $x_{s+1}, \ldots, x_{s+t}$ by least squares, and let the regression equations be

$$\xi_\alpha = \sum_i b_{\alpha i} x_i. \tag{4.7}$$

The appropriateness of Q as a generalization of the correlation coefficient, and of $\sqrt{Z}$ as a generalization of the alienation coefficient, will be apparent from the following theorem:

The ratio of the generalized variance of $\xi_1, \ldots, \xi_s$ to that of $x_1, \ldots, x_s$ is Q^2. The ratio of the generalized variance of $x_1 - \xi_1, x_2 - \xi_2, \ldots, x_s - \xi_s$ to that of $x_1, \ldots, x_s$ is Z.

This theorem is expressed in terms of the population, but an exactly parallel one holds for a sample.

(*Editors' note*: The proof has been omitted.)

A further property of the vector correlation is obvious from the final paragraph of Section 3:

A necessary and sufficient condition that the number of components in an uncorrelated set of components common to two sets of variates be less than the number of variates in either set is that the vector correlation be zero.

When $s = 2$ the canonical correlations not only determine the vector correlation and alienation coefficients but are determined by them. If as usual we take q positive, (4.6) becomes $q = r_1 r_2$, $z = (1 - r_1^2)(1 - r_2^2)$, whence

$$r_1^2 + r_2^2 = 1 - z + q^2, \qquad r_1^2 r_2^2 = q^2. \tag{4.8}$$

Solving, and denoting the greater canonical correlation by r_1, we have

$$\left.\begin{aligned} r_1 &= \tfrac{1}{2}[\sqrt{(1 + q)^2 - z} + \sqrt{(1 - q)^2 - z}] \\ r_2 &= \tfrac{1}{2}[\sqrt{(1 + q)^2 - z} - \sqrt{(1 - q)^2 - z}] \end{aligned}\right\}. \tag{4.9}$$

Since the canonical correlations are real, $(r_1 - r_2)^2$ is positive; therefore

$$z \leqslant (1 - q)^2. \tag{4.10}$$

In like manner, the vector correlation and alienation coefficients in the population are subject not only to the inequalities $0 \leqslant Q^2 \leqslant 1, 0 \leqslant Z \leqslant 1$, but also, when $s = 2$, to

$$Z \leqslant (1 - Q)^2.$$

These inequalities become equalities when the roots are equal.

The fundamental equation (3.6), regarded as an equation in λ^2, has as roots the squares of the canonical correlations. Hence, by (4.8), it reduces it to the form

$$\lambda^4 - (1 - z + q^2)\lambda^2 + q^2 = 0, \tag{4.11}$$

where $s = 2$.

5. Standard Errors

The canonical correlations and the coefficients of the canonical variates are defined in Section 3 in such a way that they are continuous functions of the covariances, with continuous derivatives of all orders, except for certain sets of values corresponding to multiple or zero roots, within the domain of variation for which the covariances are the coefficients of a positive definite quadratic form. This is true for the canonical reduction of a sample as well as for that of a population. The probability that a random sample from a continuous distribution will yield multiple roots is zero; and sample covariances must always be the coefficients of a positive definite form.

We shall in this section derive asymptotic standard errors, variances and covariances for the canonical correlations on the assumption that those in the population are unequal, and that the population has the multiple normal distribution. From these we shall derive standard errors for the vector correlation and alienation coefficients q and z. The deviation of sample from population values in these as in most cases have variances of order n^{-1}, and distributions approaching normality of form as n increases*.

Let $x_1, \ldots, x_p$ be a normally distributed set of variates of zero means and covariances

$$\sigma_{ij} = Ex_ix_j. \tag{5.1}$$

For a sample of N in which x_{if} is the value of x_i observed in the fth individual, the sample covariance of x_i and x_j is

$$s_{ij} = \frac{\sum_f (x_{if} - \bar{x}_i)(x_{jf} - \bar{x}_j)}{N - 1} = \frac{\sum x_{if}x_{jf} - N\bar{x}_i\bar{x}_j}{N - 1}, \tag{5.2}$$

* For a proof of approach to normality for a general class of statistics including those with which we deal, cf. Doob, *op. cit.*

where $\bar{x}_i$ and $\bar{x}_j$ are the sample means. To simplify the later work, we introduce the *pseudo-observations*, x'_{if}, defined in terms of the observations by the equations

$$x'_{if} = \sum_{g=1}^{N} c_{fg} x_{ig}, \tag{5.3}$$

where the quantities c_{fg}, independent of i and therefore the same for all the variates x_i, are the coefficients of an orthogonal transformation, such that

$$c_{N1} = c_{N2} = \cdots = c_{NN} = \frac{1}{\sqrt{N}}. \tag{5.4}$$

Since the transformation is othogonal we must have

$$\sum_h c_{fh} c_{gh} = \delta_{fg}, \tag{5.5}$$

where δ_{fg} is the Kronecker delta, equal to unity if $f = g$, but to zero if $f \neq g$. The coefficients c_{fg} may be chosen in an infinite variety of ways consistently with these requirements, but will be held fixed throughout the discussion. Since linear functions of normally distributed variates are normally distributed, the pseudo-observations are normally distributed. Their population means are, from (5.3),

$$Ex'_{if} = \sum c_{fg} Ex_{ig} = 0,$$

since the original variates were assumed to have zero means. Also, since the expectation of the product of independent variates is zero, and since the different individuals in a sample are assumed independent, so that, by (5.1),

$$Ex_{ih} x_{jk} = \delta_{hk} \sigma_{ij}, \tag{5.6}$$

we have, from (5.3), (5.6) and (5.5),

$$\left.\begin{aligned} Ex'_{if} x'_{jg} &= \sum_{hk}\sum c_{fh} c_{gk} Ex_{ih} x_{jk} \\ &= \sum_{hk}\sum c_{fh} c_{gk} \delta_{hk} \sigma_{ij} \\ &= \sum_h c_{fh} c_{gh} \sigma_{ij} \\ &= \delta_{fg} \sigma_{ij} \end{aligned}\right\}. \tag{5.7}$$

From (5.4) and (5.3) it is clear that

$$x'_{iN} = \frac{\sum_g x_{ig}}{\sqrt{N}} = \sqrt{N} \bar{x}_i. \tag{5.8}$$

The equations (5.3) may, on account of their orthogonality, be solved in the form

$$x_{if} = \pm \sum c_{gf} x'_{ig}.$$

Therefore, by (5.5),

$$\sum_f x_{if}x_{jf} = \sum\sum\sum c_{gf}c_{hf}x'_{ig}x'_{jh} = \sum\sum \delta_{gh}x'_{ig}x'_{jh} = \sum_g x'_{ig}x'_{jg}.$$

Substituting this result and (5.8) in (5.2), we find that the final term of the sum cancels out. Introducing therefore the symbol S for summation from 1 to $N - 1$ with respect to the second subscript, and putting also

$$n = N - 1, \tag{5.9}$$

we have the compact result

$$s_{ij} = \frac{Sx'_{ig}x'_{jg}}{n}. \tag{5.10}$$

Since the pseudo-observations are normally distributed with the covariances (5.7) and zero means, they have exactly the same distribution as the observations in a random sample of n from the original population. The equivalence of the mean product (5.10) with the sample covariance (5.2) establishes the important principle that *the distribution of covariances in a sample of* $n + 1$ *is exactly the same as the distribution of mean products in a sample of* n, if the parent population is normally distributed about zero means. Use of this principle will considerably simplify the discussions of sampling.

An important extension of this consideration lies in the use of deviations, not merely from sample means, but from regression equations based on other variates. In such cases the number of degrees of freedom n to be used is the difference between the sample number and the number of constants in each of the regression equations, which number must be the same for all the deviations. The estimate of covariance of deviations in the ith and jth variates to be used is then the sum of the products of corresponding deviations, divided by n. This may also be regarded as the mean product of the values of x_i and x_j in n pseudo-observations, as above, without elimination of the means or of the extraneous variates. The sampling distributions with which we shall be concerned will all be expressed in terms of the number of degrees of freedom n, rather than in terms of the number of observations N. This will permit immediately of the extension, which is equivalent to replacing all the correlations, in terms of which our statistics may be defined, by partial correlations representing the elimination of a particular set of variates, the same in all cases.

A variance is of course the covariance of a variate with itself, so that this whole discussion of covariances is equally applicable to variances.

The characteristic function of a multiple normal distribution with zero means is well known to be

$$M(t_1, t_2, \ldots) = Ee^{\sum t_i x_i} = e^{\sum\sum \sigma_{ij}t_it_j/2}.$$

The moments of the distribution are the derivatives of the characteristic function, evaluated for $t_1 = t_2 = \cdots = 0$. From the fourth derivative with respect to t_i, t_j, t_k and t_m it is easy to show in this way that

$$Ex_i x_j x_k x_m = \sigma_{ij}\sigma_{km} + \sigma_{im}\sigma_{jk} + \sigma_{ik}\sigma_{jm}. \tag{5.11}$$

From (5.10) we have

$$Es_{ij}s_{km} = \frac{1}{n^2} SSEx'_{ig}x'_{jg}x'_{kf}x'_{mf}. \tag{5.12}$$

Now if $f \neq g$,

$$Ex'_{ig}x'_{jg}x'_{kf}x'_{mf} = (Ex'_{ig}x'_{jg})(Ex'_{kf}x'_{mf}) = \sigma_{ij}\sigma_{km}, \tag{5.13}$$

since the expectation of the product of *independent* quantities is the product of their expectations. Of the n^2 terms in the double sum in (5.12), $n^2 - n$ are equal to (5.13). The remaining n terms are those for which $f = g$, and each of them equals (5.11). Hence

$$Es_{ij}s_{km} = \sigma_{ij}\sigma_{km} + \frac{1}{n}(\sigma_{im}\sigma_{jk} + \sigma_{ik}\sigma_{jm}).$$

Inasmuch as

$$Es_{ij} = \sigma_{ij},$$

we have, if we put

$$d\sigma_{ij} = s_{ij} - \sigma_{ij},$$

for the deviation of sample from population value, that the sampling co-variances is

$$Ed\sigma_{ij}\, d\sigma_{km} = Es_{ij}s_{km} - \sigma_{ij}\sigma_{km},$$

whence

$$Ed\sigma_{ij}\, d\sigma_{km} = \frac{1}{n}(\sigma_{ik}\sigma_{jm} + \sigma_{im}\sigma_{jk}). \tag{5.14}$$

This is a fundamental formula from which may be derived directly a number of more familiar special cases. For example, to obtain the variance of a variance, merely put $i = j = k = m$, which gives

$$\sigma_{sii}^2 = E(d\sigma_{ii})^2 = \frac{2\sigma_{ii}^2}{n}.$$

Returning from these general considerations to the problem of canonical correlations, we recall from (3.2) and (3.3) that for any particular canonical correlation ρ_1,

$$\sum\sum \sigma_{\alpha\beta}a_\alpha a_\beta = 1, \qquad \sum\sum \sigma_{ij}b_i b_j = 1, \qquad \rho_1 = \sum\sum \sigma_{\alpha i}a_\alpha b_i, \tag{5.15}$$

where α and β are summed from 1 to s, and i and j from $s+1$ to $s+t$. Any particular set of sampling errors $d\sigma_{AB}$ in the covariances determines a corresponding set of sampling errors in the a_α and b_i and in ρ_1, for these quantities are definite analytic functions of the covariances except when ρ_1 is a multiple or zero root of (3.6), cases which we now exclude from consideration. In terms of the derivatives of these functions we define

$$\begin{gathered} da_\alpha = \sum\sum \frac{\partial a_\alpha}{\partial \sigma_{AB}} d\sigma_{AB}, \qquad db_i = \sum\sum \frac{\partial b_i}{\partial \sigma_{AB}} d\sigma_{AB}, \\ d\rho_1 = \sum\sum \frac{\partial \rho_1}{\partial \sigma_{AB}} d\sigma_{AB}, \end{gathered} \tag{5.16}$$

where $d\sigma_{AB} = s_{AB} - \sigma_{AB}$, and the summations are over all values of A and B from 1 to $s+t$. Then differentiating (5.15) we have

$$\left.\begin{aligned} &\sum\sum(2\sigma_{\alpha\beta} a_\alpha da_\beta + a_\alpha a_\beta d\sigma_{\alpha\beta}) = 0, \qquad \sum\sum(2\sigma_{ij} b_i db_j + b_i b_j d\sigma_{ij}) = 0, \\ &d\rho_1 = \sum\sum(\sigma_{\alpha i} a_\alpha db_i + \sigma_{\alpha i} b_i da_\alpha + a_\alpha b_i d\sigma_{\alpha i}) \end{aligned}\right\}. \tag{5.17}$$

Let us now suppose that the variates are in the population canonical. This assumption does not entail any loss of generality as regards ρ_1, since ρ_1 is an invariant under transformations of the variates of either set. Since a_α is the coefficient of x_α in the expression for one of the canonical variates, which we take to be x_1, we have in the population $a_1 = 1, a_2 = a_3 = \cdots = a_s = 0$. In the same way,

$$b_{s+1} = 1, \qquad b_{s+2} = \cdots = b_{s+t} = 0.$$

Also, since the covariances among canonical variates are the elements of the determinant in (3.16), we have

$$\sigma_{\alpha\beta} = \delta_{\alpha\beta}, \qquad \sigma_{ij} = \delta_{ij}, \qquad \sigma_{\alpha i} = \delta_{\alpha+s,i}\rho_\alpha, \tag{5.18}$$

the Kronecker deltas being equal to unity if the two subscripts are equal, and otherwise vanishing. When these special values of the a's, b's and σ's are substituted in (5.17) most of the terms drop out, leaving the simple equations

$$\left.\begin{aligned} &2da_1 + d\sigma_{11} = 0, \qquad 2db_{s+1} + d\sigma_{s+1,s+1} = 0, \\ &d\rho_1 = \rho_1\, db_{s+1} + \rho_1\, da_1 + d\sigma_{1,s+1} \end{aligned}\right\}. \tag{5.19}$$

Substituting from the first two in the third of these equations, we get

$$d\rho_1 = d\sigma_{1,s+1} - \tfrac{1}{2}\rho_1(d\sigma_{11} + d\sigma_{s+1,s+1}). \tag{5.20}$$

For any other simple root ρ_2 we have in the same way

$$d\rho_2 = d\sigma_{2,s+2} - \tfrac{1}{2}\rho_2(d\sigma_{22} + d\sigma_{s+2,s+2}). \tag{5.21}$$

Squaring (5.20), taking the expectation, using the fundamental formula (5.14), and finally substituting the canonical values (5.18), we have

$$\left.\begin{aligned} nE(d\rho_1)^2 &= \sigma_{11}\sigma_{s+1,s+1} + \sigma_{1,s+1}^2 - \rho_1(2\sigma_{11}\sigma_{1,s+1} + 2\sigma_{s+1,s+1}\sigma_{1,s+1}) \\ &\quad + \tfrac{1}{4}\rho_1^2(2\sigma_{11}^2 + 4\sigma_{1,s+1}^2 + 2\sigma_{s+1,s+1}^2) \\ &= 1 + \rho_1^2 - \rho_1(2\rho_1 + 2\rho_1) + \tfrac{1}{4}\rho_1^2(2 + 4\rho_1^2 + 2) \\ &= (1 - \rho_1^2)^2 \end{aligned}\right\}. \tag{5.22}$$

Treating the product of (5.20) and (5.21) in the same way we obtain

$$E d\rho_1\, d\rho_2 = 0. \tag{5.23}$$

A sample canonical correlation r_1 may be expanded about ρ_1 in a Taylor series of the form

$$r_1 = \rho_1 + \sum\sum \frac{\partial \rho_1}{\partial \sigma_{AB}} d\sigma_{AB} + \tfrac{1}{2}\sum\sum\sum\sum \frac{\partial^2 \rho_1}{\partial \sigma_{AB}\partial \sigma_{CD}} d\sigma_{AB}\, d\sigma_{CD} + \cdots, \tag{5.24}$$

or, by the last of (5.16),

$$r_1 - \rho_1 = d\rho_1 + \cdots. \tag{5.25}$$

The expectation of the product of any number of the sampling deviations $d\sigma_{AB}$ is a fixed function of the σ's divided by a power of n whose exponent increases with the number of the quantities $d\sigma_{AB}$ in the product. Since $Ed\sigma_{AB} = 0$, we have from (5.24) and (5.14) that $E(r_1 - \rho_1)$ is of order n^{-1}. Hence squaring (5.25) and using (5.22), we find that the sampling variance of r_1 is given by $\dfrac{(1 - \rho_1^2)^2}{n}$, apart from terms of higher order in n^{-1}. If by the standard error of r_1 we understand the leading term in the asymptotic expansion of the square root of the variance, we have for this standard error

$$\sigma_{r_1} = \frac{1 - \rho_1^2}{\sqrt{n}}. \tag{5.26}$$

It is remarkable that this standard error of a canonical correlation is of exactly the same form as that of a product-moment correlation coefficient calculated directly from data, at least so far as the leading term is concerned.

The covariance of two statistics or their correlation would ordinarily be of order n^{-1}; but from (5.23) it appears that the covariance of r_1 and r_2 is of order n^{-2} at least. All these results hold as between any pair of simple non-vanishing roots. To summarize:

Let $\rho_1, \rho_2, \ldots, \rho_p$ be any set of simple non-vanishing roots of (3.6). *For sufficiently large samples these will be approximated by certain of the canonical correlations $r_1, r_2, \ldots, r_p$ of the samples in such a way that, when $r_\gamma - \rho_\gamma$ is divided by the standard error*

$$\sigma_{r_\gamma} = \frac{1 - \rho_\gamma^2}{\sqrt{n}} \qquad (\gamma = 1, 2, \ldots, p), \tag{5.27}$$

the resulting variates have a distribution which, as n increases, approaches the normal distribution of p independent variates of zero means and unit standard deviations.

For small samples there will be ambiguities as to which root of the determinantal equation for the sample is to be regarded as approximating a particular canonical correlation of the population. As n increases, the sample roots will separately cluster more and more definitely about individual population roots.

If a canonical correlation ρ_γ is zero, and if $s = t$, the foregoing result is applicable with the qualification that sample values r_γ approximating ρ_γ must not all be taken positive, but must be assigned positive and negative values with equal probabilities. Alternatively, if we insist on taking all the sample canonical correlations as positive, the distribution will be that of absolute values of a normally distributed variate.

To prove this, suppose that the determinantal equation has zero as a double root. For sample covariances sufficiently near those in the population, there will be a root r close to zero, which will be very near the value of λ obtained by dropping from the equation all but the term in λ^2 and that independent of λ. The latter is for $s = t$ a perfect square, and the former does not vanish, since the zero root is only a double one. Hence r is the ratio of a polynomial in the s_{AB}'s to a non-vanishing regular function in the neighbourhood. This means that the differential method applicable to non-vanishing roots is also valid here, and that, since the derivatives are continuous, (5.27) holds even when $\rho_\gamma = 0$.

Since a tetrad difference is proportional to a vector correlation, which is the product of the canonical correlations, the question whether the tetrad differs significantly from zero is equivalent to the question whether a canonical correlation is significantly different from zero. This may be tested by means of the standard error (5.27), which reduces in this case to $\frac{1}{\sqrt{n}}$. Since this is independent of unknown parameters, we have here a method of meeting the third of the difficulties mentioned in Section 1 in connection with testing the significance of the tetrad.

For $s = 2$, a zero root is of multiplicity t at least. From the final result in §9 below it may be deduced that if zero is a root of multiplicity exactly t, if r is the corresponding sample canonical correlation, and if $s = 2$, then nr^2 has the χ^2 distribution with $t - 1$ degrees of freedom. This provides a means of testing the significance of a sample canonical correlation in all cases in which $s = 2$.

We shall conclude this section by deriving standard error formulae for the vector correlation and vector alienation coefficients, assuming the canonical correlations in the population all distinct. Differentiating (4.5) and supposing all canonical correlations positive we have

$$dQ = \sum \frac{d\rho_\gamma}{\rho_\gamma}, \qquad dZ = -2Z \sum \frac{\rho_\gamma\, d\rho_\gamma}{1 - \rho_\gamma^2}.$$

Taking the expectations of the squares and products of these expressions and using (5.22) and (5.23), we obtain for the variances and covariance, apart from terms of higher order in n^{-1},

$$\sigma_q = Q\sqrt{\frac{1}{n}\sum_{\gamma=1}^{s}\frac{(1-\rho_\gamma^2)^2}{\rho_\gamma^2}}, \qquad \sigma_z = 2Z\sqrt{\frac{\rho_1^2 + \cdots + \rho_s^2}{n}},$$

$$E\, dQ\, dZ = -\frac{2}{n} QZ \sum (1 - \rho_\gamma^2). \tag{5.28}$$

For the case $s = 2$ these formulae reduce with the help of (4.5) to

$$\sigma_q = \sqrt{\frac{(1-Q^2)^2 - Z(1+Q^2)}{n}}, \qquad \sigma_z = 2Z\sqrt{\frac{1 - Z + Q^2}{n}},$$

$$E\, dQ\, dZ = -\frac{2}{n} QZ(1 + Z - Q^2).$$

6. Examples, and an Iterative Method of Solution

The correlations obtained by Truman L. Kelley* among tests in (1) reading speed, (2) reading power, (3) arithmetic speed, and (4) arithmetic power are given by the elements of the following determinant, in which the rows and columns are arranged in the order given:

$$D = \begin{vmatrix} 1.0000 & .6328 & .2412 & .0586 \\ .6328 & 1.0000 & -.0553 & .0655 \\ .2412 & -.0553 & 1.0000 & .4248 \\ .0586 & .0655 & .4248 & 1.0000 \end{vmatrix} = .4129.$$

These correlations were obtained from a sample of 140 seventh-grade school children. Let us inquire into the relations of arithmetical with reading abilities indicated by these tests.

The two-rowed minors of D in the upper left, lower right, and upper right corners are respectively

$$A = .5996, \quad B = 8195, \quad \sqrt{C} = .01904.$$

Hence, by (4.2)

$$q^2 = .0007377, \quad q = .027161, \quad z = .84036. \tag{6.1}$$

* *Op. cit.*, p. 100. These are the raw correlations, not corrected for attenuation.

By means of (4.9) or (4.11) these values give for the canonical correlations

$$r_1 = .3945, \quad r_2 = .0688. \tag{6.2}$$

In this case $n = N - 1 = 139$, and the standard error (5.27) reduces, for the hypothesis of a zero canonical correlation in the population, to $\frac{1}{\sqrt{139}} = .0848$. It is plain, therefore, that r_2 is not significant, so that we do not have any evidence here of more than one common component of reading and arithmetical abilities.

Whether we have convincing evidence of *any* common component is another question. It is tempting to compare the value of r_1 also with the standard error .0848 for the purpose of answering this question, which would give a decidedly significant value. This however is not a sensitive procedure for testing the hypothesis that there is no common factor; for this hypothesis of complete independence would mean that both canonical correlations would in the population be zero; they would therefore be a quadruple root of the fundamental equation, to which the standard error is not applicable. We conclude that reading and arithmetic involve one common mental factor but, so far as these data show, only one.

Linear functions $a_1x_1 + a_2x_2$ and $b_3x_3 + b_4x_4$ having maximum correlation with each other may be used either to predict arithmetical from reading ability or vice versa. The coefficients will satisfy (3.4) and (3.5); when in these equations we substitute $r_1 = .3945$ for λ and μ, and the given correlations for the covariances, and divide by $-\lambda = -.3945$, we have

$$\begin{aligned}
a_1 + .6328a_2 - .6114b_3 - .1485b_4 &= 0,\\
.6328a_1 + a_2 + .1402b_3 - .1660b_4 &= 0,\\
-.6114a_1 + .1402a_2 + b_3 + .4248b_4 &= 0,\\
-.1485a_1 - .1660a_2 + .4248b_3 + b_4 &= 0.
\end{aligned}$$

The fourth equation must be dependent on the preceding three, so we ignore it except for a final checking. Replacing b_4 by unity we may solve the first three equations, which are symmetrical, by the usual least-square method. Thus we write the coefficients, without repetition, in the form

1.0000	.6328	−.6114	−.1485	.8729
	1.0000	.1402	−.1660	1.6070
		1.0000	.4248	.9536

the last column consisting of the sums of the elements written or understood in the respective rows. The various divisions, multiplications and subtractions involved in solving the equations are applied to the elements in the rows, including those in the check column, which at every stage gives the sum of the elements written or understood in a row. In the array above, the coefficients of each equation begin in the first row and proceed downward to the diagonal, then across to the right, and this scheme is followed with the reduced set of equations obtained by eliminating an unknown, which is done in such a way as to preserve symmetry. This process yields finally the ratios

$$a_1 : a_2 : b_3 : b_4 = -2.772 : 2.2655 : -2.4404 : 1.$$

Therefore the linear functions of arithmetical and reading scores that predict each other most accurately are proportional to $-2.7772x_1 + 2.2655x_2$ and $-2.4404x_3 + x_4$, respectively. It is for these weighted sums that the maximum correlation .3945 is attained.

From the same individuals, Kelley obtained the correlations in the following table, in which the first two rows correspond to the arithmetic speed and power tests cited above, while the others are respectively memory for words, memory for meaningful symbols, and memory for meaningless symbols:

$$\left\| \begin{array}{cc|ccc} 1.0000 & .4248 & .0420 & .0215 & .0573 \\ .4248 & 1.0000 & .1487 & .2489 & .2843 \\ \hline .0420 & .1487 & 1.0000 & .6693 & .4662 \\ .0215 & .2489 & .6693 & 1.0000 & .6915 \\ .0573 & .2843 & .4662 & .6915 & 1.0000 \end{array} \right\| .$$

From this we find $q^2 = .0003209$, $q = .01792$, $z = .902466$, whence

$$r_1 = .3073, \qquad r_2 = .0583.$$

Since in this case $s \neq t$, we cannot say as before that the standard error of r_2 when $\rho = 0$ is $n^{-1/2} = .0848$. But, putting $\chi^2 = nr_2^2 = .472$, with two degrees of freedom, we find $P = .79$, so that r_2 is far from significant. However r_1 is decidedly significant.

In view of the tests in Section 11, we conclude in this case also that there is evidence of one common component but not of two.

If each of the two sets contains more than two variates, the two invariants q and z do not suffice to determine the coefficients of the various powers of λ in the determinantal equation, so that its roots can no longer be calculated in the foregoing manner. The coefficients in the equation will involve other rational invariants in addition to q and z, but we shall not be concerned with these, and it is desirable to have a procedure that does not require their calculation, or the explicit determination and solution of the equation. It is also desirable to avoid the explicit solution of the sets of linear equations (3.4) and (3.5) when the variates are numerous, since the labour of the direct procedure then becomes excessive. These computational difficulties are analogous to those in the determination of the principal axes of a quadric in n-space, or of the principal components of a set of statistical variates, problems for which an iterative procedure has been found useful, and has been proved to converge to the correct values in all cases*. We shall now show how

* Harold Hotelling, "Analysis of a Complex of Statistical Variables into Principal Components" in *Journal of Educational Psychology*, Vol. xxiv. pp. 417–441 and 498–520 (September and October, 1933), Section 4.

a process partly iterative in character may be applied to determine canonical variates and canonical correlations between two sets.

If in the s equations (3.4) we regard $\lambda a_1, \lambda a_2, \ldots, \lambda a_s$ as the unknowns, we may solve for them in terms of the b's by the methods appropriate for solving normal equations. Indeed, the matrix of the coefficients of the unknowns is symmetrical; and in the solving process it is only necessary to carry along, instead of a single column of right-hand members, t columns, from which the coefficients of $b_{s+1}, \ldots, b_{s+t}$ in the expressions for $a_1, \ldots, a_s$ are to be determined. The entries initially placed in these columns are of course the covariances between the two sets. Let the solution of these equations consist of the s expressions

$$\lambda a_\alpha = \sum_i g_{\alpha i} b_i \qquad (\alpha = 1, 2, \ldots, s). \tag{6.3}$$

In exactly the same way the t equations (3.5), with μ replaced by λ, may be solved for $\lambda b_{s+1}, \ldots, \lambda b_{s+t}$ in the form

$$\lambda b_i = \sum_\beta h_{i\beta} a_\beta \qquad (i = s + 1, \ldots, s + t). \tag{6.4}$$

If we substitute from (6.4) in (6.3) and set

$$k_{\alpha\beta} = \sum_i g_{\alpha i} h_{i\beta}, \tag{6.5}$$

we have

$$\lambda^2 a_\alpha = \sum_\beta k_{\alpha\beta} a_\beta. \tag{6.6}$$

Now if an arbitrarily chosen set of numbers be substituted for $a_1, \ldots, a_s$ in the right-hand members of (6.6), the sums obtained will be proportional to the numbers substituted only if they are proportional to the true values of $a_1, \ldots, a_s$. If, as will usually be the case, the proportionality does not hold, the sums obtained, multiplied or divided by any convenient constant, may be used as second approximations to solutions $a_1, \ldots, a_s$ of the equations. Substitution of these second approximations in the right-hand members of (6.6) gives third approximations which may be treated in the same way; and so on. Repetition of this process gives repeated sets of trial values, whose ratios will be seen below to approach as limits those among the true values of $a_1, \ldots, a_s$. The factor of proportionality λ^2 in (6.6) becomes r_1^2, the square of the largest canonical correlation. When the quantities $a_1', \ldots, a_s'$ eventually determined as sufficiently nearly proportional to $a_1, \ldots, a_s$ are substituted in the right-hand members of (6.4), there result quantities $b_{s+1}', \ldots, b_{s+t}'$ proportional to $b_{s+1}, \ldots, b_{s+t}$, apart from errors which may be made arbitrarily small by continuation of the iterative process. The factor of proportionality to be applied in order to obtain linear functions with unit variance is the same for the a's and the b's; from (3.2), (3.4), and (3.5) it may readily be shown that if from the quantities obtained we calculate

$$m = \frac{\sqrt{r_1}}{\sqrt{\sum\sum \sigma_{\alpha i} a'_\alpha b'_i}}, \tag{6.7}$$

then the true coefficients of the first pair of canonical variates are $ma'_1, \ldots, ma'_s, mb'_{s+1}, \ldots, mb'_{s+t}$.

In the iterative process, if $a_1, \ldots, a_s$ represent trial values at any stage, those at the next stage will be proportional to

$$a'_\alpha = \sum k_{\alpha\beta} a_\beta. \tag{6.8}$$

Another application of the process gives

$$a''_\gamma = \sum k_{\gamma\alpha} a'_\alpha,$$

whence, substituting, we have

$$a''_\gamma = \sum k^{(2)}_{\gamma\beta} a_\beta,$$

provided we put

$$k^{(2)}_{\gamma\beta} = \sum_\alpha k_{\gamma\alpha} k_{\alpha\beta}.$$

The last equation is equivalent to the statement that the matrix K^2 of the coefficients $k^{(2)}_{\gamma\beta}$ is the square of the matrix K of the $k_{\alpha\beta}$. It follows therefore that one application of the iterative process by means of the squared matrix is exactly equivalent to two successive applications with the original matrix. This means that if at the beginning we square the matrix only half the number of steps will subsequently be required for a given degree of accuracy.

The number of steps required may again be cut in half if we square K^2, for with the resulting matrix K^4 one iteration is exactly equivalent to four with the original matrix. Squaring again we obtain K^8, with which one iteration is equivalent to eight, and so on. This method of accelerating convergence is also applicable to the calculation of principal components*. It embodies the root-squaring principle of solving algebraic equations in a form specially suited to determinantal equations.

After each iteration it is advisable to divide all the trial values obtained by a particular one of them, say the first, so as to make successive values comparable. The value obtained for a_1, if this is the one used to divide the rest at each step, will approach r_1^2 if the matrix K is used in iteration, but will approach r_1^4 if K^2 is used, r_1^8 if K^4 is used, and so forth. When stationary values are reached, they may well be subjected once to iteration by means of K itself, both in order to determine r^2 without extracting a root of high order, and as a check on the matrix-squaring operations.

If our covariances are derived from a sample from a continuous multivariate distribution, it is infinitely improbable that the equation is ω,

* Another method of accelerated iterative calculation of principal components is given by T.L. Kelley in *Essential Traits of Mental Life*, Cambridge, Mass., 1935. A method similar to that given above is applied to principal components by the author in *Psychometrika*, Vol. I. No. 1 (1936).

$$\begin{vmatrix} k_{11}-\omega & k_{12} & \cdots\cdots k_{1s} \\ k_{21} & k_{22}-\omega & \cdots\cdots k_{2s} \\ \cdots & \cdots & \cdots \\ k_{s1} & k_{s2} & \cdots\cdots k_{ss}-\omega \end{vmatrix} = 0,$$

has multiple roots. If we assume that the roots $\omega_1, \omega_2, \ldots, \omega_s$ are all simple, and regard $a_1, \ldots, a_s$ as the homogeneous coordinates of a point in $s-1$ dimensions which is moved by the collineation (6.8) into a point $(a_2', \ldots, a_s')$, we know* that there exists in this space a transformed system of coordinates such that the collineation is represented in terms of them by

$$\bar{a}_1' = \omega_1 \bar{a}_1, \quad \bar{a}_2' = \omega_2 \bar{a}_2, \ldots, \bar{a}_s' = \omega_s \bar{a}_s.$$

Another iteration yields a point whose transformed homogeneous coordinates are proportional to

$$\omega_1^2 \bar{a}_1, \quad \omega_2^2 \bar{a}_2, \ldots, \omega_s^2 \bar{a}_s.$$

Continuation of this process means, if ω_1 is the root of greatest absolute value, that the ratio of the first transformed coordinates to any of the others increases in geometric progression. Consequently the moving point approaches as a limit the invariant point corresponding to this greatest root. Therefore the ratios of the trial values of $a_1, \ldots, a_s$ will approach those among the coefficients in the expression for the canonical variate corresponding to the greatest canonical correlation. Thus the iterative process is seen to converge, just as in the determination of principal components.

After the greatest canonical correlation and the corresponding canonical variates are determined, it is possible to construct a new matrix of covariances of deviations from these canonical variates. When the iterative process is applied to this new matrix, the second largest canonical correlation and the corresponding canonical variates are obtained. This procedure may be carried as far as desired to obtain additional canonical correlations and variates, as in the method of principal components; but the later stages of the process will yield results which will usually be of diminishing importance. The modification of the matrix is somewhat more complicated than in the case of principal components, and we shall omit further discussion of this extension.

(*Editors' note*: The remainder of this section, as well as sections 7–16 have been omitted.)

* Bôcher, p. 293.