First published in *Psychometrika*, 1965, *30*, 57–71.

Linear Correlations Between Sets of Variables

Abstract

While the traditional multiple correlation coefficient appears to be inherently an asymmetrical statistic, it is actually a special case of a more general measure of linear relationship between two sets of variables. Another symmetric generalization of linear correlation is to the total relatedness within a set of variables. Both of these developments rest upon the generalized variance of a multivariate distribution, which is seen to be the fundamental concept of linear correlational theory.

One of the more provocative concepts to emerge from information theory (uncertainty analysis) is that of the "total restraint" (Garner, 1962) or "total correlation" (Watanabe, 1960) within a joint distribution of categorical variables. Defined as the amount by which the total uncertainty in the distribution falls short of the maximal joint uncertainty possible for these variables given their marginal uncertainties, the total-restraint statistic is a measure of information-theoretical relationship within a set of variables which is perfectly symmetric in its arguments (i.e., it does not divide the variables whose relatedness it assesses into criterion and predictors) while if the total set of variables is partitioned into subsets, the total restraint correspondingly partitions into within-group and between-group components. Thus information theory suggests symmetric methods of relational analysis not *prima facie* available to multiple correlation theory. Moreover, if a set of variables is split dichotomously, the between-group component of the set's information-theoretical total restraint is a measure of relationship between the two subsets with precisely the same meaning as the information-theoretical relationship between two single variables. Since all recognized metrical correlational statistics accommodate but one criterion variable at a time, the ability to represent the relatedness of two *sets* of variables by a single number appears to be still another way in which information theory is capable of analytic stunts which are beyond the reach of correlation theory.

When information theory and linear correlational analysis are compared searchingly, however, it turns out that *every* information-theoretical statistic has a correlational counterpart, including the total-constraint and between-set measures of relationship. I have discussed the nature of this isomorphism in a forthcoming article (1968) which, however, makes no attempt to explore what sort of distributional significance these new correlational statistics might have. Yet worthy of exploration they most certainly are, for they appreciably deepen our understanding of linear multivariate relationships and intensify the intimacies among multiple correlation, principal components, and canonical correlation.

To describe this generalization of linear correlation theory, it is first necessary to review certain concepts and theorems that are already familiar to most students of multivariate analysis. We begin with some notational conventions. Lower-case letters $x_1, x_2, \cdots, y_1, y_2, \cdots$ etc. denote variables (variates) which have a joint distribution in some given population P. (All statistical properties cited subsequently are relative to the same parametric population P, which will henceforth be presupposed without explicit mention.) Capital letters X, Y, etc. will denote ordered *sets* of variables, not excluding the null set. When we wish to make explicit what variables are included in a set X we will write $X = \langle x_1, \cdots, x_n \rangle$ or the like, where $n \geq 0$. To deal with the case $n = 0$ without requiring additional assumptions or definitions, we may posit that every set of variables $X = \langle x_1, \cdots, x_n \rangle$ also includes an additional variable x_0 whose value is the same for every member of P. The reason for describing the set $X = \langle x_1, \cdots, x_n \rangle$ as *ordered* is only to avoid the unnecessary restriction that $x_i \neq x_j$ if $i \neq j$. Actually, the variables within each set X, Y, etc. in the equations below are freely permutable, as will be obvious in all cases except the one for which proof is supplied (Theorem 4).

Let *the space, S_X, of variables spanned by the set of variables $X = \langle x_1, \cdots, x_n \rangle$* be defined as the set of all variables which are linear combinations of the variables in X. That is, $y \in S_X$ if and only if there exist constants $a_1, \cdots, a_n$ such that

$$y = a_o + \sum_{i=1}^{n} a_i x_i \,.$$

Any set of variables $\langle x'_1, \cdots, x'_m \rangle$, none of which is a linear combination of the others and which are such that $y \in S_X$ if and only if y is a linear combination of $x'_1, \cdots, x'_m$, is said to be a *basis* for space S_X. A basis $X' = \langle x'_1, \cdots, x'_m \rangle$ for space S_X is an *orthogonal* basis if the covariances among $x'_1, \cdots, x'_m$ are all zero, while if in addition each x'_i in X' has zero mean and unit variance, X' is an *orthonormal* basis for S_X. Any space S_X of variables has at least one orthonormal basis—in fact, if the dimensionality of S_X is greater than unity, it has an infinitude of them. (The dimensionality of a space S_X is the number of variables in any basis for S_X, and is also the rank of the covariance matrix for any set of variables which span S_X.) By definition, a variable y is *orthogonal* to a space S_X, i.e., $y \perp S_X$, if Var $(y) > 0$ and Cov $(y, x) = 0$ for every $x \in S_X$, while two spaces S_X and S_Y are orthogonal if every variable in the one has zero covariance with every variable in the other. If Var $(y) > 0$, a necessary and sufficient condition for $y \perp S_X$ is for y to be orthogonal to (i.e., have zero covariance with) all the variables in some basis for S_X.

We shall write $\dot{y}_{(X)}$ for the linear regression of a variable y upon variables

$X = \langle x_1, \cdots, x_n \rangle$, while $e_{y \cdot X}$ is the residual of y unaccounted for by $x_1, \cdots, x_n$. Then

$$(1) \qquad y = \dot{y}_{(X)} + e_{y \cdot X} \ ,$$

where $\dot{y} \in S_X$ and $e_{y \cdot X} \perp S_X$. A fundamental theorem of linear regression theory is that if X' is any other set of variables which also spans S_X, then

$$\dot{y}_{(X')} = \dot{y}_{(X)} \quad \text{and} \quad e_{y \cdot X'} = e_{y \cdot X} \ .$$

Thus the (linear) partition of a criterion variable into regressed and residual components is a function only of the space spanned by the predictor variables and not, in addition, of the particular *way* in which predictor space is spanned. We shall here call the component $\dot{y}_{(X)}$ the *projection* of variable y into space S_X, though strictly speaking it is $\dot{y}_{(X)} - M_y$ (where M_y is the mean of y) which is most properly regarded as y's projection into S_X. Also, we shall write $\sigma_{y\,(X)}$ and $\sigma_{y \cdot X}$ for the standard deviations of $\dot{y}_{(X)}$ and $e_{y \cdot X}$ respectively. The quantity $\sigma^2_{y \cdot X}$ is known as the residual variance of y after variables $X = \langle x_1, \cdots, x_n \rangle$ have been partialled out. In the limiting case when X is empty (or equivalently, when X contains only the constant-variable x_o), $\sigma_{y \cdot X} = \sigma_y$.

The multiple correlation of a variable y with the variables in a set X may be defined

$$(2) \qquad R_{y\,(X)} =_{\text{def}} \frac{\sigma_{y\,(X)}}{\sigma_y}.$$

In the vector model of a multivariate configuration, $R_{y\,(X)}$ is the cosine of the angle between Y and space S_X (i.e., between Y and y's projection into S_X). Analytically, $R_{y\,(X)}$ is the linear correlation between y and $\dot{y}_{(X)}$ and is the maximal linear correlation between y and any variable in S_X. Similarly, the multiple coefficient of alienation, $K_{y\,(X)}$, for the predictability of variable y from variables X is

$$(3) \qquad K_{y\,(X)} =_{\text{def}} \frac{\sigma_{y \cdot X}}{\sigma_y}.$$

which in the vector model is the sine of y's angle to S_X . Since $e_{y \cdot X}$ is orthogonal to $\dot{y}_{(X)}$,

$$(4) \qquad \sigma^2_y = \sigma^2_{y\,(X)} + \sigma^2_{y \cdot X} \ ,$$

or, dividing by σ^2_y,

$$(5) \qquad R^2_{y(X)} + K^2_{y(X)} = 1.$$

(The expression "vector model" used in the preceding paragraph is somewhat ambiguous, since there are actually two major versions of this. The first, which is well defined both for frequency distributions and for probability distributions and might be called the "factorial vector model," represents each variable by a vector whose coordinates are the variable's factor coefficients on some orthonormal basis for a space which includes all the variables under concern. In this model, a variable's standard deviation is equal to the length of its representative vector while the linear correlation between two variables equals the cosine of the angle between the corresponding vectors. When what is being modeled is a joint distribution of frequencies in a population containing N members, however, it is also possible to represent each variable x by a vector in N-space whose coordinates are the x-scores of the various population members. The latter model is especially useful in sampling theory and might hence be called the "sampling vector model." Standard deviations and correlations are also represented in the sampling vector model, but not so simply as in the factorial vector model. Whenever in this paper the ambiguity in the term "vector model" makes a difference, it is the factorial vector model which is intended.)

Given a vector space S of variables, we can partition all variables in which we are interested into projections into S and residuals orthogonal to S, and proceed to examine the relations which are found within either set of components. The multiple correlation of the y-residual $e_{y\,\cdot\,Z}$ with the x_i-residuals $e_{x_1\,\cdot\,Z}, \cdots, e_{x_n\,\cdot\,Z}$ orthogonal to a space S_Z spanned by a set of variables Z is known as a multiple-partial correlation and will here be written $R_{y\,(X)\,\cdot\,Z}$. Similarly, the multiple-partial coefficient of alienation $K_{y\,(X)\,\cdot\,Z}$ is the multiple coefficient of alienation for predicting $e_{y\,\cdot\,X}$ from the $e_{x_i\,\cdot\,Z}$ Since the component of $e_{y\,\cdot\,Z}$ orthogonal to the space spanned by residual set $E_{X\,\cdot\,Z}$ ($=_{\text{def}}$ $e_{x_1\,\cdot\,Z}, \cdots, e_{x_n\,\cdot\,Z}$) is identical with the component of y orthogonal to the spaces spanned by the combined predictors in $X = \langle x_1, \cdots, x_n \rangle$ and $Z = \langle z_1, \cdots, z_n \rangle$ (another fundamental regression theorem), we may write

(6) $$e_{(e_y\,\cdot\,Z)\,\cdot\,E_{X\,\cdot\,Z}} = e_{y\,\cdot\,XZ}\ ,$$

where $XZ = \langle x_1, \cdots, x_n, z_1, \cdots, z_m \rangle$. Hence

(7) $$K_{y\,(X)\,\cdot\,Z} = \frac{\sigma_{y\,\cdot\,XZ}}{\sigma_{y\,\cdot\,Z}}\ ,$$

(8) $$R_{y\,(X)\,\cdot\,Z} = \sqrt{1 - K^2_{y\,(X)\,\cdot\,Z}} = \sqrt{1 - \left(\frac{\sigma_{y\,\cdot\,XZ}}{\sigma_{y\,\cdot\,Z}}\right)^2}\ .$$

In conformity to prevailing custom, the multivariate-alienation statistic K may be written with a lower-case k when there is only one predictor, i.e., the alienation between two variables z and y after the variables in set Z have been partialled out is

(9) $$k_{xy\,\cdot\,Z} = \frac{\sigma_{y\,\cdot\,xZ}}{\sigma_{y\,\cdot\,Z}} = \sqrt{1 - r^2_{xy\,\cdot\,Z}}\ ,$$

where $r_{xy \cdot Z}$ is standard notation for partial correlation.

It is frequently convenient in multivariate analysis to transform a given set of variables X into another set X' which is equivalent to the first for the purpose at hand. In descriptive factor analysis, for example, it is customary to reduce the original variables to an orthonormal basis for the space they span. Another possibility is to replace the set X with an *orthogonal rotation* thereof. A set of variables $X' = \langle x'_1, \cdots, x'_n \rangle$ is an orthogonal rotation of set $X = \langle x_1, \cdots, x_n \rangle$ when there exists an orthogonal $n \times n$ matrix $\{a_{ij}\}$ of constants such that for $i = 1, \cdots, n$,

$$x'_i = \sum_{i=1}^{n} a_{ij} x_i \ .$$

(By definition, an $n \times n$ matrix M is orthogonal when $MM' = I$, where I is the $n \times n$ identity matrix.) In a *scatter model* (i.e., scattergram or probability-density surface) for the joint distribution of variables $X = \langle x_1, \cdots, x_n \rangle$, wherein the n variables are represented by n mutually perpendicular Cartesian coordinate axes to define a scatter space within which each member of population P is assigned a position by his joint scores on the x_i, an orthogonal rotation of X is an alternative set of orthogonal coordinate axes for this same scatter space. An important property of orthogonal rotations is that if a set of variables X' is an orthogonal rotation of set X, then X' not merely spans the same space of variables as does X but also the total variance of the variables in X' accounted for by any further set of variables Y is the same as the total variance of the variables in X accounted for by Y. That is, if X' is an orthogonal rotation of X, then

$$\sum_{i=1}^{n} \sigma^2_{x'_i(Y)} = \sum_{i=1}^{n} \sigma^2_{x_i(Y)}$$

Since the scatter model of a multivariate distribution is *not* the same as the vector model thereof, it is not generally the case that the variables in an orthogonal rotation of set X are orthogonal in the vector-model sense of zero covariances.

There is, however, one very special basis for the space spanned by variables X which manages to embrace the configurational ideals of both the vector and the scatter models simultaneously, namely, the *principal components* of the X distribution. Specifically, the principal components of variables X are an orthogonal rotation of X in which the rotated variables are *also* orthogonal in the vector model. Thus the principal components of the multivariate configuration X are both an orthogonal rotation of X (hence preserving the total variance of the set lying in any given predictor space) and an orthogonal (though not in general orthonormal) basis for the vector space S_X spanned by set X. (More precisely, if there are more variables in X than there are dimensions to S_X, the principal components of X with nonzero variance are a basis for S_X.) It is conventional to number the principal components of a set of variables in decreasing order of their variances, i.e., if $\sigma^2_{(X)_i}$ is the variance, or *eigenvalue*, of the ith principal component of set X, then $\sigma^2_{(X)_1} \geq \sigma^2_{(X)_2} \geq \cdots \geq \sigma^2_{(X)_n}$. A further basic property of principal components is that the total variance of set X accounted for by its ith principal component is equal to $\sigma^2_{(X)_i}$, while this is also the largest amount of the total residual variance in set X which can be accounted for by any one variable after the preceding $i-1$ principal components have been partialled out. Thus described wholly in terms of the vector model, the principal components of set X are an orthogonal basis for space S_X such that each successive basis-variable maximizes the amount of total X variance it accounts for while its own variance is equal to this accounted-for variance. It can also be shown that the principal-component variances $\sigma^2_{(X)_1}, \cdots, \sigma^2_{(X)_n}$ are the latent roots of the covariance matrix C_{XX} (i.e., $\{\text{Cov}(x_i, x_i)\}$) for the variables in X. The product,

$$\prod_{i=1}^{n} \sigma^2_{(X)_i} ,$$

of these principal-component variances is equal to the determinant of the variance matrix C_{XX} and is known as the *generalized variance* of the multivariate distribution X (see Anderson, 1958, p. 167). Thus the generalized variance $|C_{XX}|$ of set X is a scalar measure of the total n-dimensional dispersion of population P on the variables in X and subsumes the variance of a single variable as the special case in which $n = 1$. We shall have more to say about this little-known statistic shortly.

Principal components provide a standardized description of the variance structure *within* a set of variables. Similarly, *canonical factors* and *canonical correlations* are a standardized description of the variance relations which hold between two sets of variables. The canonical factors of sets $X = \langle x_1, \cdots, x_n \rangle$ and $Y = \langle y_1, \cdots, y_m \rangle$ with respect to each other are two series of linear combinations of $x_1, \cdots, x_n$ and $y_1, \cdots, x_m$, respectively, such that the ith canonical factor of X and the ith canonical factor of Y (with respect to each other) are the two variables lying in S_X and S_Y, respectively, which have maximal linear correlation with each other subject to the restriction that they are orthogonal to all the $i-1$ preceding canonical factors. The ith canonical correlation $r_{(X,Y)_i}$ between sets X and Y is, of course, the linear correlation between their ith canonical factors. Since the definition of canonical factors as just given determines them only up to a linear transformation, it is conventional to stipulate that their means are set at zero and their variances are standardized to unity. (This is still insufficient to guarantee absolute uniqueness, but it does not seem necessary to go more deeply into the matter here.)

Canonical factors have a number of interesting properties, the most important of which will be listed here with brief proofs.

THEOREM 1. *The canonical factors of variables* $X = \langle x_1, \cdots, x_n \rangle$ *with respect to variables* $Y = \langle y_1, \cdots, y_n \rangle$ *are a sequence of mutually orthogonal linear combinations* $w_{(X)_1}, w_{(X)_2}, \cdots$ *of variables X such that the multiple correlation of* $w_{(X)_i}$ *with variables Y is maximal, subject to the prior successive maximization of the multiple correlations of* $w_{(X)_1}, \cdots, w_{(X)_{i-1}}$ *with variables* Y; *and similarly for the canonical factors of Y with respect to X. The* ith *canonical factor of Y(X) with respect to X(Y) coincides up to a linear transformation with the projection into* $S_Y(S_X)$ *of the* ith *canonical factor of X(Y) with respect to Y(X).*

PROOF. (By induction on i.) Since the multiple correlation of any given linear combination $w_{(X)}$ of variables X with variables Y is the maximal correlation that $w_{(X)}$ can have with any variable in S_Y, and similarly with X and Y interchanged, the theorem is obviously true for $i = 1$. For the induction step, assume the theorem to be true for the first $i-1$ canonical factors and consider the correlation between linear composites $w_{(X)}$ and $w_{(Y)}$ of X and Y, respectively, where $w_{(X)}$ is orthogonal to all the first $i-1$ canonical factors $w_{(X)_1}, \cdots, w_{(X)_{i-1}}$ of X with respect to Y. w_Y can be analyzed as its projection into the subspace S'_Y of S_Y spanned by the first $i-1$ canonical factors $w_{(Y)_1}, \cdots, w_{(Y)_{i-1}}$ of Y with respect to X, plus a residual orthogonal to S'_Y. But the component of $w_{(Y)}$ lying in this subspace S'_Y is orthogonal to $w_{(X)}$, for it is a linear composite of the first $n-1$ canonical factors of Y with respect to X, and these in turn analyze into their projections into S_X—which by the induction hypothesis respectively coincide up to a linear transformation with the corresponding canonical factors of X and are hence all orthogonal to $w_{(X)}$—plus residuals which are orthogonal to space S_X and hence *a fortiori* to $w_{(X)}$. Consequently, the component of $w_{(Y)}$ orthogonal to $w_{(Y)_1}, \cdots, w_{(Y)_{i-1}}$ has greater correlation with $w_{(X)}$ than does $w_{(Y)}$ unless the latter is itself wholly orthogonal to these factors; and hence the projection of $w_{(X)}$ into S_Y is also orthogonal to $w_{(Y)_1}, \cdots, w_{(Y)_{i-1}}$. In particular, if $w_{(X)}$ is a variable in S_X whose multiple correlation with the variables in Y is greatest of all the variables in S_X orthogonal to the first $i-1$ canonical factors of X with respect to Y, while $w_{(Y)}$ is the projection of $w_{(X)}$ into S_Y, it follows that $w_{(X)}$ and $w_{(Y)}$—or, more precisely, linear standardizations thereof to zero means and unit variances—qualify as the ith canonical factors of sets X and Y, respectively, with respect to each other. QED.

It may be noted that the wording of Theorem 1 evades commitment to how many canonical factors sets X and Y have with respect to each other. Unlike the definition of principal components, in which it is inherent that a set of n variables has exactly n principal components, the definition of canonical correlations leaves their number to some extent conventional. The maximum number is the dimensionality of either S_X or S_Y, depending on which is smaller, and this is the convention usually adopted; however, if each of the spaces S_X or S_Y contains a subspace which is orthogonal to the other, one or more of the terminal pairs of canonical factors will be arbitrary, corresponding to zero-valued canonical correlations. An alternative, more symmetric convention would be to stipulate that the number of canonical factors is equal to the number of nonzero canonical correlations. In either case, it is clear from Theorem 1 and the stipulation that canonical factors have zero means and unit variances that the m canonical factors of set X with respect to set Y can always be chosen as the first m variables in an orthonormal basis for space S_X. It is also obvious from either Theorem 1 or the initial definition of canonical factors that

THEOREM 2. *The canonical factors and canonical correlations between two sets of variables S_X and S_Y are determined wholly by the spaces S_X and S_Y spanned by sets X and Y, respectively.*

That is, if sets X and X' span the same space S_X and sets Y and Y' span the same space S_Y, the canonical factors of X and Y with respect to each other are identical with the canonical factors of X' and Y' with respect to each other. Consequently, canonical factors and canonical correlations are most insightfully described as a set of relational properties between the spaces spanned by the two sets of variables in question.

THEOREM 3. *The* ith *canonical factor* ($i = 1, 2, \ldots$) *of a space of variables S_Y with respect to another space of variables S_X coincides up to a linear transformation with the* ith *principal component of the set of projections into S_Y of any orthonormal basis for space S_X, while the standard deviation of this principal component equals the* ith *canonical correlation between S_X and S_Y.*

PROOF. Let $X = \langle x_1, \cdots, x_n \rangle$ be an orthonormal basis for space S_X. Then for any variable y, it is elementary to show that

$$R^2_{y(X)} = \sum_{i=1}^{n} r_{yx_i} = \sum_{i=1}^{n} \sigma^2_{x_i(y)} \ ,$$

which says that the squared multiple correlation of a variable y with any set of variables spanning a space S_X equals the total variance of an orthonormal basis for S_X accounted for by y. Moreover, if $y \in S_Y$, the variance of x_i accounted for by y equals the variance of x_i's projection into S_Y accounted for by y (since $\sigma^2_{x_i(y)}$ equals the variance of $\dot{x}_{(Y)}$ accounted for by y plus the variance of $e_{x_i \cdot Y}$ accounted for by y, while if $y \in S_Y$ the latter is zero). Hence for any variable y in space S_Y, the squared multiple correlation of y with any set of variables spanning space S_X is equal to the total variance of the projections into S_Y of an orthonormal basis for S_X accounted for by y—of which Theorem 3 is then a simple consequence in light of Theorem 1.

We are now properly positioned to appreciate the new correlational measures disclosed by the isomorphism between linear correlational analysis and information theory. Let the quantity Π_X for the joint distribution of variables X be defined

$$\begin{aligned} \Pi_X &=_{\text{def}} \prod_{i=1}^{n} \sigma_{x_i \cdot x_1 \cdots x_{i-1}} \quad (X = \langle x_1, \cdots, x_n \rangle) \\ &= \left[\prod_{i=1}^{n} K_{x_i(x_1 \cdots x_{i-1})} \right] \cdot \left[\prod_{i=1}^{n} \sigma_{x_i} \right] \end{aligned} \tag{10}$$

Elsewhere (Rozeboom, 1968), I have unimaginatively named Π_X the *Pi-value* of multivariate distribution X. However, in view of Theorem 5, below, it may also be appropriately called the "generalized standard deviation" of set X.

THEOREM 4. *Π_X is invariant under permutations of the variables in X.*

PROOF. Since

$$\sigma_{x \cdot Z} \sigma_{y \cdot xZ} = \sigma_{x \cdot Z} \sigma_{y \cdot Z} k_{xy \cdot Z} = \sigma_{y \cdot Z} \sigma_{x \cdot yZ}$$

for any two variables x, y and additional variables Z, any two adjacent variables x_1 and x_{i+1} in definition (10) may be interchanged without affecting the value of Pi_X—which by iteration proves the theorem.

THEOREM 5. *Π_X is the square root of the generalized variance of the set of variables X. That is,*

$$\Pi_X =_{\text{def}} \prod_{i=1}^{n} \sigma_{(X)_i} \quad (X = \langle x_1, \cdots, x_n \rangle), \tag{11}$$

where $\sigma_{(X)_i}$, is the standard deviation of the ith *principal component of variables X.*

PROOF. It is easily seen in any of a number of ways that the quantity $\sigma_{x_i} \sigma_{x_j \cdot x_i}$ for any two variables x_i and x_j is unaffected by orthogonal rotation of the pair. (E.g., the determinant of the covariance matrix for these two variables is

$$\begin{vmatrix} V_i & C_{ij} \\ C_{ij} & V_j \end{vmatrix} = V_i V_j - C_{ij}^2 = \sigma_{x_i}^2 \sigma_{x_j}^2 (1 - r_{x_i x_j}^2) = \sigma_{x_i}^2 \sigma_{x_j \cdot x_i}^2 \;\;,$$

while it is well known that the determinant of a covariance matrix is indifferent to orthogonal rotations of the variables.) Hence in view of Theorem 4, Π_X is unaffected by orthogonal rotation of any pair of variables in X; and since any orthogonal rotation of a finite number of variables can be accomplished by a finite number of pairwise rotations, Π_X is accordingly invariant under all orthogonal rotations of set X. But the principal components of X are an orthogonal rotation of X, so their Pi-value is equal to Π_X, while since principal components are also uncorrelated, their Pi-value is equal to the product of their standard deviations. QED.

THEOREM 6. *In the (factorial) vector model of a multivariate distribution,* Π_X *equals the volume of the n-dimensional parallelotope whose principal vertex and edges are formed by the configuration of vectors which represent the variables in X.*

Multidimensional geometric forms are difficult to visualize, and even more so to describe in nonmathematical terms, so rather than attempting to clarify the meaning of this theorem, I refer the reader to Anderson (1958, p. 167 f.), who develops an equivalent of Theorem 6 for the sampling vector model. However, Anderson's matrix-algebra proof of this theorem is much less intuitive than what is possible using definition (10) and Theorem 5. When $n = 2$, the geometric figure envisioned by Theorem 6 is the parallelogram of which the vectors representing variables x_1 and x_2 form adjacent edges. The lengths of these edges are σ_{x_1} and σ_{x_2}, respectively, while the area of a parallelogram is the length of one edge, i.e., σ_{x_1}, times the length of the projection of the other edge orthogonal to the first, i.e., $\sigma_{x_2 \cdot x_1}$. Each additional dimension multiplies the volume of the preceding $(n-1)$-dimensional side by the length of the nth edge's projection orthogonal to the other edges (i.e., $\sigma_{x_n \cdot x_1 \cdots x_{n-1}}$), which by induction on n establishes the theorem.

Theorem 6 is mentioned here only to convey some feeling for the sense in which Π_X is a measure of multidimensional spread. In the vector model, Π_X may be construed as the volume of space occupied by the configuration of vectors representing X. Π_X also has a volume interpretation in the scatter model which, though it will not be discussed here, can readily be visualized from Theorem 5 by considering the ellipsoidal volume of scatter-space occupied by an n-variate normal distribution whose covariance matrix is the same as that of variables X.

In obvious application of the generic operation of "partialling" in multivariate analysis, the residual Pi-value, $\Pi_{X \cdot Z}$, of variables $X = (x_1, \cdots, x_n)$ after variables $X = (z_1, \cdots, z_m)$ have been partialled out is the Pi-value of the set of components of $x_1, \cdots, x_n$ orthogonal to space S_Z. From (6) and (10) we then have

$$\text{(12)} \qquad \begin{aligned} \Pi_{X \cdot Z} &= \prod_{i=1}^{n} \sigma_{x_i \cdot x_1 \cdots x_{i-1} Z} \quad (X = \langle x_1, \cdots, x_n \rangle) \\ &= \frac{\Pi_{XZ}}{\Pi_Z} \, , \end{aligned}$$

where $XZ = \langle x_1, \cdots, x_n, z_1, \cdots, z_m \rangle$ and the equivalence between $\Pi_{X \cdot Z}$ and Π_{XZ}/Π_Z presupposes that none of the variables in Z is a linear function of the others (since otherwise both Π_{XZ} and Π_Z become zero even though $\Pi_{X \cdot Z}$ remains well defined).

It will be observed in the second line of (10) that Π_X can be analyzed as a product of two terms, one of which, $\prod_{i=1}^{n} \sigma_{x_i}$ is a function wholly of the (generally arbitrary) units of measurement by which the variables in X are scaled, while the other is determined entirely by the correlations (or, more immediately, by the alienations) among the variables in X and is unaffected by any linear rescalings of the variables. Accordingly, let the *Pi-coefficient* π_X for a configuration of variables X be defined

(13)
$$\begin{aligned} \pi_X &=_{\text{def}} \frac{\Pi_X}{\prod_{i=1}^{n} \sigma_{x_i}} \quad (X = \langle x_1, \cdots, x_n \rangle) \\ &= \prod_{i=1}^{n} K_{x_i(x_1 \cdots x_{i-1})} \, . \end{aligned}$$

The statistic π_X might also be called the "generalized alienation coefficient" or better, *within-set alienation*, for it is a standardized measure of the lack of inter-predictability among the variables in X and reduces to the familiar coefficient of alienation $k_{x_1 x_2}$ in the special case in which $n = 2$. The corresponding measure of *within-set correlation* would be

(14)
$$\rho_X =_{\text{def}} \sqrt{1 - \pi_X^2} \, ,$$

though the only reason for preferring ρ_X to π_X as a measure of the over-all relatedness in X is that ρ_X increases as the linear relationships within X grow tighter, whereas π_X correspondingly decreases. Both ρ_X and π_X are bounded by 0 and 1, with $\rho_X = 0$ and $\pi_X = 1$ when the variables in X are all orthogonal to each other, while the opposite extreme $\rho_X = 1$ and $\pi_X = 0$ is reached when at least one of the variables in X is an errorless linear function of the others. In passing, it may be noted that ρ_X and π_X are *not* invariant under orthogonal rotation of X (as is readily appreciated by considering that the within-set correlation for the principal components of X is necessarily zero), and that the Pi-coefficient for the residuals of variables X after variables Z have been partialled out is

(15)
$$\begin{aligned} \pi_{X \cdot Z} &=_{\text{def}} \frac{\Pi_{X \cdot Z}}{\prod_{i=1}^{n} \sigma_{x_i \cdot Z}} \quad (X = \langle x_1, \cdots, x_n \rangle) \\ &= \frac{\Pi_{X \cdot Z}}{\prod_{i=1}^{n} \Pi_{x_i \cdot Z}} \\ &= (n-1) \frac{\Pi_{XZ} \Pi_Z}{\prod_{i=1}^{n} \Pi_{x_i \cdot Z}} \, , \end{aligned}$$

the last line of which follows by application of general principle (12) under the condition that set Z contains no linear dependencies.

As just seen, the familiar concept of the linear correlation between two variables has a natural extension to a symmetric measure of linear relatedness among n variables. An even more interesting extension of two-variable correlation is to the linear correlation between two sets of variables. Assuming Π_X and Π_Y to be nonzero (a condition which will later be dropped), let the *between-set alienation* $K_{X,Y}$ be defined

(16)
$$\begin{aligned} K_{X,Y} &=_{\text{def}} \frac{\Pi_{XY}}{\Pi_X \Pi_Y} \\ &= \frac{\pi_{XY}}{\pi_X \pi_Y} \end{aligned}$$

(the second line of which follows by application of (13)), while the corresponding *between-set correlation* is

(17)
$$\begin{aligned} R_{X,Y} &=_{\text{def}} \sqrt{1 - K_{X,Y}^2} \\ &= \sqrt{1 - \left(\frac{\pi_{XY}}{\pi_X \pi_Y}\right)^2} \,. \end{aligned}$$

To show that $K_{X,Y}$ and $R_{X,Y}$ are indeed a straightforward generalization of traditional correlation measures, we observe that in the special case where one of the sets, say Y, contains only one variable y,

(18)
$$K_{y,X} = \frac{\sigma_{y \cdot X}}{\sigma_y} = K_{y\,(X)} \,,$$

while similarly when $Y = y$, $R_{Y,X}$ reduces to the multiple correlation of y with the variables in X. More generally, from (16) and (12),

(19)
$$K_{X,Y} = \frac{\Pi_{X \cdot Y}}{\Pi_X} = \frac{\Pi_{X \cdot Y}}{\Pi_Y} \,,$$

in which the formal similarity to (3) is conspicuous. The alienation between sets X and Y after the variables in a third set Z have been partialled out is

(20)
$$\begin{aligned} K_{X,Y \cdot Z} &= \frac{\Pi_{XY \cdot Z}}{\Pi_{X \cdot Z} \Pi_{Y \cdot Z}} \\ &= \frac{\Pi_{X \cdot YZ}}{\Pi_{X \cdot Z}} = \frac{\Pi_{Y \cdot XZ}}{\Pi_{Y \cdot Z}} \\ &= \frac{\Pi_{XYZ} \Pi_Z}{\Pi_{XZ} \Pi_{YZ}} \,, \end{aligned}$$

The corresponding partial between-set correlation is of course

$$R_{X,Y \cdot Z} = \sqrt{1 - K^2_{X,Y \cdot Z}} \ ,$$

of which the multiple-partial correlation coefficient $R_{y(X) \cdot Z}$ is the special case in which $Y = y$.

Since $\Pi_{Y \cdot X}$ is the Pi-value of the set of Y-components orthogonal to the space spanned by variables X, it is unaffected by replacing X with any other set of variables which span S_X. Consequently, in view of (19),

THEOREM 7. *The correlation $R_{X,Y}$ and alienation $K_{X,Y}$ between two sets of variables X and Y are unchanged by replacing either X or Y with another set of variables which spans the same space as the set replaced.*

Hence $R_{X,Y}$ and $K_{X,Y}$ are fundamentally measures of relationship between the spaces S_X and S_Y. Theorem 7 also shows that the definition of $K_{X,Y}$, which in (16) presupposes sets with nonzero Pi-coefficients, can be released from this restriction. Specifically, if either $\Pi_X = 0$ or $\Pi_Y = 0$, $K_{X,Y}$ is stipulated to equal $K_{X',Y'}$, where X' and Y' are any bases for the spaces spanned by X and Y, respectively.

Together, Theorems 2 and 7 suggest that a connection should exist between canonical correlations and between-set correlation. This is indeed the case, and delightfully so:

THEOREM 8. *Let $r_{(X,Y)_i}$ be the* ith *canonical correlation between two sets of variables X and Y, while*

$$k_{(X,Y)_i} = \sqrt{1 - r^2_{(X,Y)_i}}$$

is the corresponding ith *canonical alienation coefficient. Then*

$$K_{X,Y} = \prod_{i=1}^{m} k_{(X,Y)_i} \ , \qquad (21)$$

or, equivalently,

$$R_{X,Y} = \sqrt{1 - \prod_{i=1}^{m} (1 - r^2_{(X,Y)_i})}, \qquad (22)$$

where m is the number of nonzero canonical correlations between sets X and Y.

PROOF. Let n be the dimensionality of the space S_X spanned by set X, where of course $n \geq m$. In view of Theorem 1, we can always find an orthonormal basis $X' = \langle x'_1, \cdots x'_n \rangle$ for S_X in which the first m variables are the canonical factors of X with respect to Y. Each variable x'_i in X' may be analyzed into its projection $\dot{x}'_{i(Y)}$ into space S_Y (where $\dot{x}'_{i(Y)}$ has zero variance if $i > m$) plus its residual $e_{x'_i \cdot Y}$

orthogonal to S_Y, where in view of the orthogonalities and projection behavior described in Theorem 1,

$$\mathrm{Cov}(e_{x'_i \cdot Y},\ e_{x'_j \cdot Y}) = \mathrm{Cov}(e_{x'_i \cdot Y} + \dot{x}_{i(Y)},\ e_{x'_j \cdot Y} + x'_{j(Y)}) = \mathrm{Cov}(x'_i,\ x'_j) = 0$$

when $i \neq j$. Insomuch as the residuals $e_{x'_i \cdot Y} (i = 1, \ldots n)$ are thus all mutually orthogonal, $\Pi_{X' \cdot Y} = \prod_{i=1}^{n} \sigma_{x_i \cdot Y}$, while $\Pi_{x'} = 1$ since variables X' have been stipulated to be orthonormal. But $x'_1, \cdots x'_m$ are the canonical factors of X with respect to Y, so

$$\sigma_{x'_i \cdot Y} = \sigma_{x'_i} \sqrt{1 - r^2_{(X,Y)_i}} = k_{(X,Y)_i}$$

for $i = 1, \cdots m$, while $\sigma_{x'_i \cdot Y} = \sigma_{x'_i} = 1$ for $i = m+1, \cdots, n$. Hence

$$K_{X,Y} = K_{X',Y} = \Pi_{X' \cdot Y} \,/\, \Pi_{X'} = \sum_{i=1}^{m} k_{(X,Y)i} \qquad \text{QED}$$

It is an obvious consequence of (16) that

$$\Pi_{XY} = \Pi_X K_{X,Y} \Pi_Y, \tag{23}$$

$$\pi_{XY} = \pi_X K_{X,Y} \pi_Y. \tag{24}$$

If $K_{X,Y}$ had no interpretive significance beyond its definitional introduction in (16) as an abbreviation for the ratio $\Pi_{XY}/\Pi_X\Pi_Y$, (23) and (24) would be trivial. However, it is not (16) but (22) which reveals the fundamental nature of between-set alienation (and, correspondingly, of between-set correlation), just as it is (11), rather than definition (10), which makes clear the significance of Π_X, and as a result, (23) and (24) are enticingly informative. Thus (23) is equivalent to

$$\prod_{i=1}^{n+p} \sigma_{(XY)_i} = \left[\prod_{i=1}^{n} \sigma_{(X)_i}\right] \left[\prod_{i=1}^{m} k_{(X,Y)_i}\right] \left[\prod_{i=1}^{p} \sigma_{(Y)_i}\right] \tag{25}$$

(where n and p are the number of variables in sets X and Y, respectively), in which a property of the principal components of a multivariate configuration is multiplicatively decomposed into properties of the principal components and canonical factors of subsets of the variables, and which under iteration can partition a generalized standard deviation (i.e., Pi-value) into a product of canonical alienation coefficients in an enormously large number of different ways. Similarly, (24) states that when a set of variables is dichotomously partitioned into subsets, the total within-set alienation correspondingly partitions as a product of within-subset and between-subset alienations.

In fact, while the most conspicuous virtue of between-set correlation (or better, between-set alienation) lies in the powerful elegance with which this concept fuses the various strands of past developments in linear correlation theory into a more comprehensive unity, there is at least one data-analysis prospect which leaps out from (23) and (24). One of the perennial problems in multivariate analysis is how to search within an aggregate of variables for smaller-sized clusters within which the homogeneity is substantially greater than the similarity between clusters. This is in effect asking how the over-all correlational agreement within the total set of variables can be concentrated within subgroups at the expense of between-group agreement. But in view of (24), this is precisely how one would also describe the operation of dividing the variables into two subsets between which the alienation is maximal or, equivalently, whose between-set correlation is minimal. Each subset so identified can similarly be dichotomized into maximally alienated sub-subsets and so on until either the clusters which remain have sufficient internal homogeneity to satisfy the investigator's desire, or no possibility of further partitioning remains. (In this context, the homogeneity of a group of variables would most appropriately be measured by the minimum between-set correlation that can be produced by some dichotomous partition of the group.) What is especially interesting about this procedure is that it determines a hierarchical group-structure which is a function only of the *spaces* spanned by the various possible combinations of the variables included in the analysis, irrespective of how thickly or sparsely the variables populate these spaces. The computational demands of multivariate analysis based on between-set correlations are severe even for a high-speed computer, perhaps so beyond the limits of practicality. Even so, the method is sufficiently distinctive to merit an investigation of whether it may not perhaps point to significant structural properties of multivariate data which are overlooked by more familiar correlational techniques.

References

Anderson, T. W. (1958). *Introduction to multivariate statistical analysis.* New York: Wiley.

Garner, W. R. (1962). *Uncertainty and structure as psychological concepts.* New York: Wiley.

Rozeboom, W. W. (1968). The theory of abstract partials—an introduction. *Psychometrika*, *33*, 133–167.

Watanabe, S. (1960). Information theoretical analysis of multivariate correlation. *IBM Journal of Research and Development*, *33*, 66–82.