Dominance, Information, and Hierarchical Scaling of Variance Space

David J. Krus and Robert W. Ceurvorst
Arizona State University

A method for computation of dominance relations and for construction of their corresponding hierarchical structures is presented. It is shown that variance can be computed from the squared pairwise differences between scores and that dominance indices are actually linear transformations of variances. The interpretation of variance as a quantitative measure of information is suggested and conceptual partition of variance into components associated with relational spaces is proposed. The link between dominance and variance allows integration of the mathematical theory of information with least squares statistical procedures without recourse to logarithmic transformations of the data.

Over 30 years ago, Truman Lee Kelley (1947) suggested a method for computing variance from the squared differences between each value and every other value in a set of data. Antedating the formal definition of information in terms of change by Shannon and Weaver (1949), Kelley stressed that

> ... it is axiomatic that if all measures of a series are and must be alike there is nothing to investigate or report. The primary phenomenon of a series is that there are differences between the measures. The concept of variability is primary and that of central tendency secondary. (p. 199)

However, Kelley's assertion of the primacy of variability over the measures of central tendency went unheeded, possibly for didactic and computational reasons. Since the concept of a mean is easier to comprehend than the concept of variance, and since the total distance of points in a test space from their mean is a minimum, the variance remained conceptualized as a second moment about the mean. This view, computationally at least, is the most parsimonious, leading to an interpretation of variance in geometric terms as the average size of squared areas located in the first two quadrants about the mean, as in Figure 1. This classical interpretation of variance, where "deviations themselves are represented as linear distances from a common reference point and squared deviations (are) represented by areas, namely squares" was provided inter alios by Guilford and Fruchter (1978, p. 67).

On the other hand, when variance is computed from the pairwise differences among the scores, it can be represented most easily as the total amount of information contained in a space of the n^2 order, as is shown below. This conceptualization of variance as a measure of information fits well with earlier attempts to link the mathematical theory of information with standard least squares statistical methods (cf. Garner & McGill, 1956). A major advantage of the approach taken in this paper is that the link is made without recourse to $\log_2$ transformations of the data.

From this perspective, conceptually simple procedures are developed for the identification of hierarchical structures (i.e., dominance relations) among subjects and among variables represented in a data matrix. It is also shown how a subjects × variables analysis of variance solution can be obtained from the dominance indices.

Hierarchical Structure of Data Vectors

Consider a vector $\mathbf{x}$ of n test scores. A major difference matrix $\tilde{\mathbf{D}}$ is defined as

$$\tilde{\mathbf{D}} = \mathbf{x} \sqcap \mathbf{x}' = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} \sqcap [x_1 \ x_2 \ \ldots \ x_n] \tag{1}$$

$$= \begin{bmatrix} (x_1 - x_1) & (x_1 - x_2) & \ldots & (x_1 - x_n) \\ (x_2 - x_1) & (x_2 - x_2) & \ldots & (x_2 - x_n) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ (x_n - x_1) & (x_n - x_2) & \ldots & (x_n - x_n) \end{bmatrix} .$$

Since the elements of $\tilde{\mathbf{D}}$ are symmetric, but with opposing signs along the zero-filled principal diagonal, the squaring of each element d_{ij} would render the resulting matrix symmetric and thus 50% redundant. To eliminate this redundancy (i.e., in effect, to utilize each pairwise difference only once), all negative elements in $\tilde{\mathbf{D}}$ are set equal to zero. This operation can be symbolized as

$$\mathbf{D} = \mathbf{x} \sqsupset \mathbf{x}' \tag{2}$$

which is identical to the operation in Equation 1 except for the restriction that if $x_i < x_j$, then $(x_i - x_j)$ is set equal to zero. If the elements of $\mathbf{x}$ are arranged in ascending or descending order, $\mathbf{D}$ is a triangular matrix, i.e.,

$$D = x \daleth x' = \begin{bmatrix} 0 & 0 & \ldots & 0 \\ (x_2 - x_1) & 0 & \ldots & 0 \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ (x_n - x_1) & (x_n - x_2) & \ldots & 0 \end{bmatrix}. \qquad [3]$$
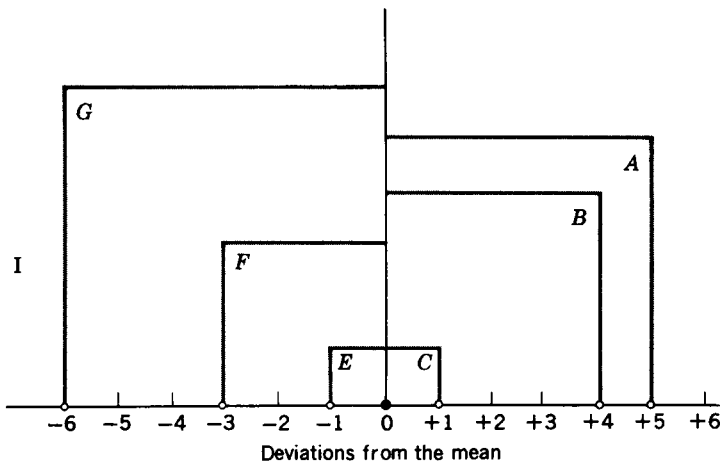
If a matrix **S** is defined, where $s_{ij} = d_{ij}^2$, then the maximum likelihood variance of **x** can be written as

$$\sigma_x^2 = \frac{1'S1}{n^2} \qquad [4]$$

where **1** is a column vector of unities and **1**′ is its transpose. Using summation notation, Equation 4 is equivalent to

$$\sigma_x^2 = \frac{\sum\limits_{i=1}^{n} \sum\limits_{j \geq i}^{n} (x_i - x_j)^2}{n^2}, \qquad [5]$$

Figure 1

The Concept of Variance as an Area Adjacent to Deviation Scores

Deviations from the mean

which can be proven to equal the more common variance formula

$$\sigma_x^2 = \frac{\sum\limits_{i=1}^{n} x_i^2}{n} - \left(\frac{\sum\limits_{i=1}^{n} x_i}{n}\right)^2 \tag{6}$$

as follows. Expansion of Equation 5 yields

$$\sigma_x^2 = \frac{(n-1)\sum\limits_{i=1}^{n} x_i^2 - 2\sum\limits_{i=1}^{n}\sum\limits_{j \geq i}^{n} x_i x_j}{n^2} . \tag{7}$$

Adding the quantity $(\sum\limits_{i=1}^{n} x_i)^2/n^2$ to both sides gives

$$\sigma_x^2 + \left(\frac{\sum\limits_{i=1}^{n} x_i}{n}\right)^2 = \frac{(n-1)\sum\limits_{i=1}^{n} x_i^2 - 2\sum\limits_{i=1}^{n}\sum\limits_{j \geq i}^{n} x_i x_j + (\sum\limits_{i=1}^{n} x_i)^2}{n^2} . \tag{8}$$

But since

$$(\sum\limits_{i=1}^{n} x_i)^2 = \sum\limits_{i=1}^{n} x_i^2 + 2\sum\limits_{i=1}^{n}\sum\limits_{j \geq i}^{n} x_i x_j , \tag{9}$$

the cross-product terms in Equation 8 cancel, leaving

$$\sigma_x^2 + \left(\frac{\sum\limits_{i=1}^{n} x_i}{n}\right)^2 = \frac{n \sum\limits_{i=1}^{n} x_i^2}{n^2} = \frac{\sum\limits_{i=1}^{n} x_i^2}{n} . \tag{10}$$

and finally,

$$\sigma_x^2 = \frac{\sum\limits_{i=1}^{n} x_i^2}{n} - \left(\frac{\sum\limits_{i=1}^{n} x_i}{n}\right)^2 \tag{11}$$

as was intended to be proven.

Example

As an example, consider a vector of scores $x' = [1\ 2\ 3\ 4\ 5]$ with a mean of 3 and a variance, computed from Equation 6, equal to $55/5 - (15/5)^2 = 2$. The matrix D is

$$
\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{bmatrix} \neg [1\ 2\ 3\ 4\ 5] \;=\; \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 \\ 3 & 2 & 1 & 0 & 0 \\ 4 & 3 & 2 & 1 & 0 \end{bmatrix}
\tag{12}
$$

which contains information about the directional differences (or dominance relations) among subjects' responses and further implies the hierarchical structure shown in Table 1.

The matrix S is obtained by squaring the elements of D with the result

$$
S = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 4 & 1 & 0 & 0 & 0 \\ 9 & 4 & 1 & 0 & 0 \\ 16 & 9 & 4 & 1 & 0 \end{bmatrix} .
\tag{13}
$$

Then, using Equation 4, the variance is

$$
\frac{1\,'S\,1}{n^2} = \frac{50}{5^2} = 2,
\tag{14}
$$

Table 1
Hierarchical Structure of the
Sample Vector $x = [12345]$

Subjects	Intersubject distances	Cumulative distances	Dominance scale
e			4
|	1	4	
d			3
|	1	3	
c			2
|	1	2	
b			1
|	1	1	
a			0

which, of course, is the same value as obtained from Equation 6. Note that in Equation 4 the variance is defined as the sum of the squared intersubject differences divided by the dimensionality of the space (5 × 5). Since each nonzero element of **S** corresponds to a number of unit differences (or changes), their sum indexes the total amount of information (or unit changes) in the data. Hence, the concept of variance as an information measure is clearly illustrated. Additional interpretations of variance computed from dominance relations for binary data were presented by Cliff (1977) and Krus (1976).

Hierarchical Structures of Data Matrices

In this section, the procedures presented above are generalized to matrices containing responses of subjects on two or more variables (or test items, as the case may be). It is shown that dominance relations among subjects and among items can be used to generate a two-way analysis of variance solution.

Consider a data matrix **X** with m rows and k columns with elements x_{ij}, $i=1, 2, \ldots, m; j=1, 2, \ldots, k$. A sample matrix is presented in Table 2 with $m=5$ and $k=5$.

The total dominance or information content in **X**, ignoring the row and column structure, can be computed using Equations 2 and 4 with $n = mk$, i.e., by collapsing all mk elements of **X** into a single vector. For the matrix in Table 2, $\mathbf{D}_{total}$ and $\mathbf{S}_{total}$ are 25 × 25 matrices. The sum of the elements of $\mathbf{S}_{total}$ is 2,944; therefore the total variance in **X** is $2,944/25^2 = 4.71$. The number 2,944 represents the total number of unit changes in **X** or, in other words, the total dominance.

Dominance indices can also be computed for rows and columns of the matrix. For rows, an $m \times m$ matrix **P** is constructed in a manner analogous to **D** with elements $p_{ii'} = \sum_j (x_{ij} - x_{i'j})$ subject to the same restriction as before, namely that if $x_{ij} < x_{i'j}$, then $(x_{ij} - x_{i'j})$ is set equal to zero. The matrix **P** for the sample data is contained in Table 3. The p_{14} element, which represents the degree to which Subject (row) a dominates Subject (row) d, for example, is computed as

$$p_{14} = (9-3) + (7-4) + (6-5) + (5-6) + (4-8) = 10,$$

because the last two terms are negative and therefore set equal to zero. Similarly,

$$p_{41} = (3-9) + (4-7) + (5-6) + (6-5) + (8-4) = 5,$$

Table 2
Sample Data Matrix

Subjects	Items				
	A	B	C	D	E
a	9	7	6	5	4
b	7	5	4	3	2
c	4	3	2	1	0
d	3	4	5	6	8
e	2	3	4	5	7

Table 3

Matrix **P** Containing Dominance Relations

Among Rows of the Data Matrix

	a	b	c	d	e
a	0	10	21	10	13
b	0	0	11	5	7
c	0	0	0	1	2
d	5	10	17	0	5
e	3	7	13	0	0

since the first three terms, being negative, are set to zero. The other elements of **P** are obtained in the same manner. The matrix **P** contains information about the directional differences among the rows of **X** and will be a triangular matrix if there is a unidimensional order among the subjects (as in Table 1), i.e., if for every pair of subjects the direction of dominance is the same on every variable, or else zero.

From **P**, an $m \times m$ matrix **Q** is constructed in which the elements $q_{ii'}$ contain the "net" dominance of row i over row i'. The elements of **Q** are $q_{ii'} = p_{ii'} - p_{i'i}$ if $p_{ii'} > p_{i'i}$ and zero otherwise. The matrix **Q** for the sample data is presented in Table 4.

For example,

$$q_{14} = p_{14} - p_{41} = 10 - 5 = 5$$

and

$$q_{41} = p_{41} - p_{14} = 5 - 10 = -5\,(0),$$

where the last term is set equal to zero because the difference is negative.

A dendogram adjacent to the matrix **Q** showing the hierarchical relations among the rows of **X** is shown in Figure 2. A line was drawn between each node (a,b,. . .,e) if their mutual distance was not

Table 4

Matrix Q Containing Net Dominances

Among Rows of the Data Matrix

	a	b	c	d	e
a	0	10	21	5	10
b	0	0	11	0	0
c	0	0	0	0	0
d	0	5	16	0	5
e	0	0	11	0	0

Figure 2
Hierarchical Structure Among Subjects for the Sample Data
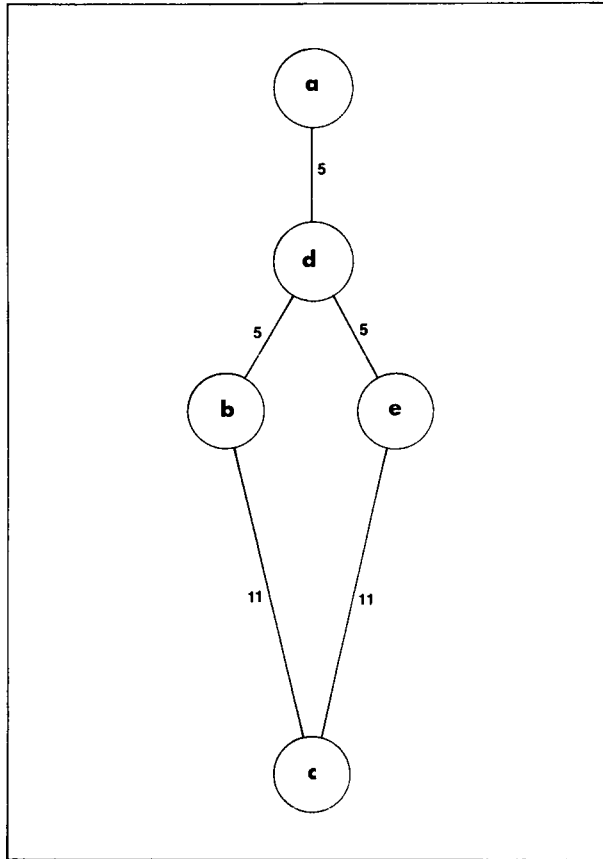
zero, as indicated by the corresponding elements of Table 4. The minimal number of lines connecting nodes of the dendogram was drawn. Thus, for example, the distance of 10 units between Nodes a and b was not plotted directly, since there is a nonzero distance between Nodes a and d and d and b of 5 units each corresponding to the 10-unit distance between Nodes a and b (cf. Table 4 and Figure 2).

Analogous operations can be performed on the columns of **X**. Thus, a $k \times k$ matrix **G** is computed, containing the directional dominance relations among the columns (items), as in Table 5. For example,

$$g_{14} = (9-5) + (7-3) + (4-1) + (3-6) + (2-5) = 11,$$

since the last two terms are negative and therefore set equal to zero. Thus, 11 is the number of units by which Item (column) 1 dominates Item 4. Similarly,

$$g_{41} = (5-9) + (3-7) + (1-4) + (6-3) + (5-2) = 6,$$

since the first three terms are set equal to zero. Here, 6 is the number of units by which Item 4 dominates Item 1.

Table 5

Matrix **G** Containing Dominance Relations

Among Columns of the Data Matrix

	A	B	C	D	E
A	0	5	8	11	14
B	2	0	3	6	9
C	4	2	0	3	6
D	6	4	2	0	3
E	10	8	6	4	0

Then, a $k \times k$ matrix **H** containing the net dominance is constructed in a manner analogous to that used to obtain **Q**. Thus, $h_{jj'} = g_{jj'} - g_{j'j}$ if $g_{jj'} > g_{j'j}$ and zero otherwise. The matrix **H** for the sample data is shown in Table 6. In this matrix,

$$h_{14} = g_{14} - g_{41} = 11 - 6 = 5$$

and

$$h_{41} = g_{41} - g_{14} = 6 - 11 = -5 \, (0),$$

where, as before, the negative differences are set equal to zero. A dendogram corresponding to **H** is displayed in Figure 3.
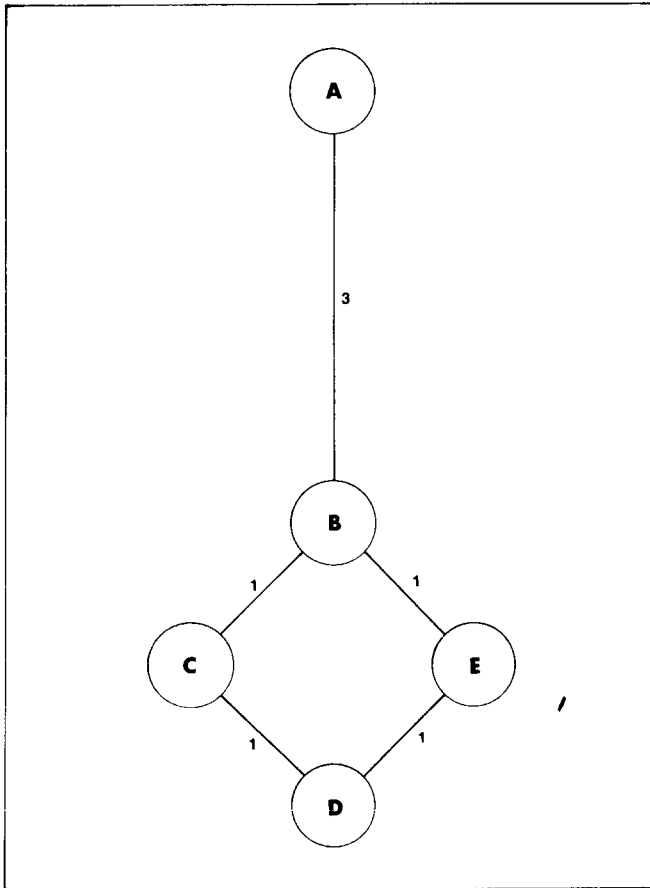
The **S** matrices for rows and columns of the matrices $\mathbf{S}_r$ and $\mathbf{S}_{c'}$ contain the squared elements of the matrices **Q** and **H**, respectively, as shown in Tables 7 and 8.

Some modifications of Equation 4 are necessary in order to obtain the row and column variances (mean squares), just as would be necessary with the traditional variance formula in Equation 6. These adjustments are required because of the summation over rows or columns in the construction of the **P**

Table 6

Matrix **H** Containing Net Dominances

Among Columns of the Data Matrix

	A	B	C	D	E
A	0	3	4	5	4
B	0	0	1	2	1
C	0	0	0	1	0
D	0	0	0	0	0
E	0	0	0	1	0

Figure 3
Hierarchical Structure of Items for the Sample Data

and **G** matrices (and thus **Q** and **H** also). Specifically, the unbiased variances (MS) for rows and columns are given by

$$\sigma_r^2 = \frac{1 \, ' \, S_r \, 1}{km(m\text{-}1)} \tag{15}$$

$$\sigma_c^2 = \frac{1 \, ' \, S_c \, 1}{km(k\text{-}1)} \tag{16}$$

where m is the number of rows; and k the number of columns. The reader may verify that application of Equation 15 to a column vector (or of Equation 16 to a row vector) also yields unbiased variance estimates, whereas Equation 4 yields the maximum likelihood variance.

Table 7

Matrix S_r for Rows of the Data Matrix

	a	b	c	d	e
a	0	100	441	25	100
b	0	0	121	0	0
c	0	0	0	0	0
d	0	25	256	0	25
e	0	0	121	0	0

Note: General element $s_{ij} = q_{ij}^2$.

The dominance indices and the corresponding sums of squares from an ordinary analysis of variance are displayed in Table 9. The dominance for the residual was obtained by subtraction (2,944−1,214−74=1,656). The proportions of variance were computed by dividing each dominance index by the total dominance; but they could just as easily have been obtained by performing the same operations with the sums of squares, with the same results. Thus, the relation of dominance to traditional sums of squares and mean squares is well illustrated.

Discussion

In the experience of the authors, the concept of variance as a measure of information may be presented and discussed without recourse to matrix algebra (by focusing on the sums of squares formula for variance in Equation 5) and thus employed in introductory statistics classes. This conceptualization of variance and of the process of partitioning variance as a formal quantitative procedure for defining meaning in the data seems to be more motivating than the alternative view of variance as the

Table 8

Matrix S_c for Columns of the Data Matrix

	A	B	C	D	E
A	0	9	16	25	16
B	0	0	1	4	1
C	0	0	0	1	0
D	0	0	0	0	0
E	0	0	0	1	0

Note: General Element $s_{ij} = h_{ij}^2$.

Table 9

Analysis of Variance Results for the Sample Data

Source	df	Sum of Squares	Dominance	Proportion of Variance	F	P
Rows	4	48.56	1214	.41236	2.932	.054
Columns	4	2.96	74	.02514	.179	.946
Residual	16	66.24	1656	.56250		
Total	24	117.76	2944	1.00000		

second moment about the mean. Unlike earlier attempts to interface the mathematical theory of information with statistical theory (cf. Attneave, 1959; Diamond, 1959; Garner & McGill, 1956), the present approach does not require $\log_2$ transformations of the observations and thus stays closer to the standard conceptual formulations of least squares statistical theory.

When the procedures described in this paper are applied to vectors, the resulting dominance relations will always imply a hierarchical structure with a single order (as in Table 1). In the case of matrices, multiple orders may be present, as in the example used here. Further discussion of the problem of multiple orders was presented by Krus and Bart (1974) and by Krus (1977) for the nonmetric binary formulation of the discussed topics, treated under the name of ordering theory and order analysis.

The theoretical and practical applications of the suggested metric model of order analysis are manifold. Elementary concepts involved in the analysis of variance can be well explicated using the dominance-variance approach, and insight into the relationship between analysis of variance and methods of hierarchical scaling may be gained. The method may be used within the context of Thurstonian scaling and, in any case, where more detailed information about sources of variation within unidimensional measuring instruments is needed. Also, Krus (1979) has recently described a method of computing test reliability (Cronbach's alpha) from dominance relations. Thus, it seems that the suggested relationships have wide applicability. Further theoretical work within the discussed context appears promising for bringing forth new insights regarding many commonly used statistical and measurement procedures.

References

Attneave, I. *Applications of information theory to psychology*. New York: Holt, 1959.

Cliff, N. A theory of consistency of ordering generalizable to tailored testing. *Psychometrika*, 1977, *42*, 375-399.

Diamond, S. *Information and error*. New York: Basic Books, 1959.

Garner, W. R., & McGill, W. J. The relation between information and variance analyses. *Psychometrika*, 1956, *21*, 219-228.

Guilford, J. P., & Fruchter, B. *Fundamental statistics in psychology and education* (6th ed.). New York: McGraw-Hill, 1978.

Kelley, T. L. *Fundamentals of statistics*. Cambridge,