Journal of Classification 6:233-246 (1989)

A Classification of Presence/Absence Based Dissimilarity Coefficients

F. B. Baulieu

University of Wisconsin

Abstract: Several desirable order properties for dissimilarity coefficients based on presence/absence of attributes are given and several popular dissimilarity coefficients are examined with respect to these properties. A characterization for rational functions with linear numerator and linear denominator satisfying all of the desirable properties is given.

Keywords: Measures of association; Profile similarities.

1. Introduction

Considerable literature is available concerning dissimilarity measures based on presence/absence of attributes. The majority of these works develop their dissimilarity coefficients heuristically. Unfortunately, there is great disagreement over the heuristics employed. In this paper we try to limit the choices for construction of dissimilarity coefficients by characterizing those which satisfy a short list of axioms. Before beginning, we make several assumptions and definitions.

We are given a finite set, S, of objects, and a finite set, A, of n binary attributes, $n > 0$. Every object, x, is associated with a mapping $v_x \in \{0,1\}^A$.

The author would like to acknowledge the meticulous corrections and helpful suggestions, including an improvement on my proof of Proposition 2, provided by the referees.

Author's Address: F. B. Baulieu, Natural and Applied Sciences, University of Wisconsin, Green Bay, WI 54301, USA.

Thus v_x may be thought of as a binary vector with one "slot" for each attribute in A. Indeed, v_x is defined by

$$v_x(a) = \begin{cases} 1 & \text{if } x \text{ exhibits attribute } a \\ 0 & \text{if } x \text{ lacks attribute } a. \end{cases}$$

Thus, given any attribute, if v_x contains a 1 in the "slot" belonging to that attribute, then object x possesses (or exhibits) that attribute. Similarly, a 0 indicates that x lacks the corresponding attribute.

Definition. A *dissimilarity coefficient* (DC) is a mapping,

$$D{:}S \times S \to \boldsymbol{R}^+$$

such that $D(x,x) = 0 \; \forall \, x \in S$, and $D(x,y) = D(y,x) \; \forall \, x, y \in S$.

Given any two binary n-vectors, v_x and v_y, it is usual to define four dependent variables as follows:

a = the number of attributes which x and y share
b = the number of attributes exhibited by x but lacked by y
c = the number of attributes exhibited by y but lacked by x
d = the number of attributes exhibited by neither x nor y.

Many dissimilarity coefficients are written as functions of a, b, c, and d alone rather than as functions of the original objects x and y. For this reason, we will define a *Presence/Absence Based Dissimilarity Coefficient* (PABDC) to be a map

$$D{:}\, (\mathbb{Z}^+)^4 - \{(0,0,0,0)\} \to \boldsymbol{R}^+$$

from the set of all ordered 4-tuples of non-negative integers other than the origin into the non-negative reals, which satisfies properties (P0)-(P7) below.

For all $(a,b,c,d) \in (\mathbb{Z}^+)^4 - \{(0,0,0,0)\}$:

(P0) D is defined for all four-tuples, (a,b,c,d) of non-negative integers except, perhaps, (0,0,0,0)
(P1) $0 \le D(a,b,c,d) \le 1$ for all a,b,c,d
(P2) $D(a,0,0,0) = 0$ for all $a > 0$
(P3) $D(a,b,c,d) = 1$ for some (a,b,c,d)
(P4) $D(a+1,b,c,d) \le D(a,b,c,d)$ with equality holding iff $D(a,b,c,d) = 0$

(P5) $D(a,b+1,c,d) \geq D(a,b,c,d)$ with equality holding iff $D(a,b,c,d) = 1$

(P6) $D(a,b,c,d+1) \leq D(a,b,c,d)$

(P7) $D(a,b,c,d) = D(a,c,b,d)$.

2. Discussion

When we decide to work with PABDCs rather than DCs, we automatically restrict the amount and type of information our coefficients can use to determine dissimilarity between objects. For instance, since we only see a, b, c, and d, we no longer know *which* attributes were present in both objects under examination. No single attribute may be singled out as "more significant" than any other. And, of course, no quantitative attributes can be directly considered.

However, working with PABDCs provides more mathematical structure to work with, and permits a reasonable avenue for classification of PABDCs with various properties. The interpretation and reasoning behind the properties (P0) to (P7) are given below. It should be noted that although some of these axioms might appear to be of marginal importance (P0) or excessively weak (P6) when dealing with a particular data set, it is the author's intent to provide the most general framework possible — structure which will be applied uniformly to *all* data sets. The conclusions reached herein may be strengthened considerably for restricted data sets with particular interpretations imposed.

In (P0) we require that the coefficient be defined for all possible 4-tuples. Since the sum $a+b+c+d$ is equal to the number of attributes in the attribute set, A, we can safely assume that (0,0,0,0) is not in the domain, but we can rule out no other 4-tuple of non-negative integers.

In (P1) we further require that the coefficient produce values in the interval [0,1]. Most coefficients in use satisfy this restraint, and those which do not can often be adjusted linearly to do so, making it easier to compare two PABDCs.

The definition of a DC required that it have a reflexive property. However, since we are using PABDCs, we no longer have the option of examining the objects themselves. We want to say, then, that if there are no "mismatches" among the attributes for two objects then the objects are indistinguishable. We provide a weaker requirement in (P2) which will, along with the other requirements, produce the full reflexive property. All we need at this point is to require that if two objects share every attribute in the attribute set, then they are identical; i.e., they have a coefficient of 0.

We require in (P3) that it be possible to artificially produce a maximum dissimilarity value by careful selection of a 4-tuple, (a,b,c,d). Again this is a condition commonly satisfied by the PABDCs currently used. If a pair of objects is imagined as disagreeing on every attribute then it is reasonable to expect the objects to be completely dissimilar; i.e., $D(x,y)$ should be 1.

In (P4) we require that when the attribute set, A, is augmented by an additional attribute, and two objects share that attribute, then the dissimilarity between those objects must decrease, if possible. After all, if the two objects were not "identical" then we are now increasing the proportion of attributes on which they agree, and the assigned dissimilarity value should be reduced. If, however, the two objects were already judged to be "identical," then this new evidence is only confirmation, and carries no additional information; in such a case, the dissimilarity value should be unchanged — at zero.

We continue this line of reasoning with a similar condition governing new attributes added to A on which two objects disagree. If the two objects were already judged maximally dissimilar, then again, this new attribute is only confirmation, and carries no additional information. Otherwise, this new attribute increases the proportion of attributes on which the two objects disagree, and thus the dissimilarity value for the pair should increase. This is the content of (P5).

The next condition, (P6), addresses an issue on which the users of cluster analysis disagree: what should be done with d? The conflict has to do with the interpretation of "matching absences." When two objects jointly lack an attribute, does that make the objects more similar? There are some cases where a binary attribute represents a dichotomy in which the dominant or significant state is not clear. That is to say, it is not clear which of the two states is to be considered "presence" and which is to be "absence." One could ask if a white object exhibits whiteness or lacks blackness. In such cases, we would want both states to be treated with equal weight. On the other hand, a quality is often specific enough that there are so many ways that an "absent" state could develop that this attribute carries little information when absent in two objects, and need not imply at all that they are similar. But neither do such "negative matches" imply that the objects are *dis*similar. It is reasonable in both interpretations, however, to assume that an increase in the number of joint absences for a pair of objects should *not* cause an *increase* in the dissimilarity measure between the objects. It is this fairly weak assumption which is included in our axioms.

Finally, as we are assuming that our PABDC depends only on the ordered 4-tuple (a,b,c,d) rather than the associated object pair, we have no information about the attributes or their distribution of 1's and 0's, or which attributes produce matches/mismatches. In that light, a mismatch of one type is equivalent to a mismatch of the other type, and b and c should be interchangeable. Thus (P7), requiring that the PABDC be *symmetric* in b and c.

3. Some Immediate Results

Several properties can be derived from (P0) through (P6). We list three of them here:

(P9) $D(0,b,c,0) = 1 \ \forall\, b\, c \in \mathbb{Z}^+, b + c > 0$
(P10) $D(a+1,b,c,d) < D(a,b+1,c,d) \quad \forall\, a,b,c,d \in \mathbb{Z}^+$
(P11) $D(a,0,0,d) = 0 \quad \forall\, a,d \in \mathbb{Z}^+,\ a + d > 0.$

The proofs are straightforward: (P9) follows from (P3) and repeated applications of (P4) and (P6). (P10) follows from (P4) and (P5). And we have (P11) since (P1), (P6), and (P2) produce the inequality

$$0 \le D(a,0,0,d) \le D(a,0,0,0) = 0\ .$$

These properties are easily interpreted, and can be seen to be appropriate: (P11) is a necessary and sufficient condition for a PABDC to satisfy the reflexive property explicitly required for a DC, namely that an object is identical to itself. (P10) simply indicates that switching a "match" to a "mismatch" in the attribute comparisons for a pair of objects must increase the dissimilarity measure for the pair. And, of course, any pair of objects which disagree on every attribute will be considered "completely dissimilar" and will have a dissimilarity of 1, as claimed in (P9).

4. One Class of PABDC

We now turn out attention to those PABDCs which satisfy (P0) through (P7) and one additional requirement:

(P8) D is a rational function whose numerator and denominator are both (total) linear. Thus D may be written in the form

$$D(a,b,c,d) = \frac{\alpha' a + \beta' b + \gamma' c + \delta' d + \varepsilon'}{\alpha a + \beta b + \gamma c + \delta d + \varepsilon}$$

for fixed $\alpha,\beta,\gamma,\delta,\varepsilon,\alpha',\beta',\gamma',\delta',\varepsilon' \in \boldsymbol{R}$.

Proposition 1. *D is a PABDC satisfying (P0) through (P8) iff D can be written in the form*

$$D(a,b,c,d) = \frac{b + c}{\alpha a + b + c + \delta d}$$

where $\alpha,\delta > 0$.

Proof. It is a simple exercise to verify that any PABDC written in the given form will satisfy (P0) through (P8). To show the converse, we begin with the expression for D provided by (P8),

$$D(a,b,c,d) = \frac{\alpha' a + \beta' b + \gamma' c + \delta' d + \varepsilon'}{\alpha a + \beta b + \gamma c + \delta d + \varepsilon}$$
$$\text{for fixed } \alpha,\beta,\gamma,\delta,\varepsilon,\alpha',\beta',\gamma',\delta',\varepsilon' \in \boldsymbol{R}\,,$$

and proceed to identify the values of those fixed coefficients. Clearly, (P11) immediately forces $\alpha' = \delta' = \varepsilon' = \mathbf{0}$, and (P0) forces at least one of α or ε to be non-zero.

Next, consider (P7) when $d = 0$. Resolving denominators in this equation we find that

$$(\beta' b + \gamma' c)(\alpha a + \beta c + \gamma b + \varepsilon) = (\beta' c + \gamma' b)(\alpha a + \beta b + \gamma c + \varepsilon)$$
$$\forall\, a,b,c \in \mathbb{Z}^+\,,\ (a,b,c) \neq (0,0,0)\,,$$

which reduces to

$$\alpha(\beta' - \gamma')ab + \alpha(\gamma' - \beta')ac + (\gamma'\beta - \beta'\gamma)c^2 +$$
$$(\gamma\beta' - \beta\gamma')b^2 + \varepsilon(\beta' - \gamma')b + \varepsilon(\gamma' - \beta')c = 0$$
$$\forall\, a,b,c \in \mathbb{Z}^+\,,\ (a,b,c) \neq (0,0,0)\,,$$

which is a polynomial in three independent variables, a, b, and c, whence all of its coefficients must be 0. In particular, using the fact that not both α and ε are zero, we get $\beta' = \gamma'$. We therefore know that $\beta' = \gamma' \neq 0$ by (P9). Returning our attention to the coefficients in the polynomial above, then, the coefficient of the third term now produces $\beta = \gamma$. Thus

$$D(a,b,c,d) = \frac{\beta'(b + c)}{\alpha a + \beta(b + c) + \delta d + \varepsilon}$$

and now (P9) forces $(\beta - \beta')(b + c) + \varepsilon = \mathbf{0}$ where b and c are independent variables, so that we can conclude that $\varepsilon = \mathbf{0}$ and $\beta = \beta'$. Notice also that (P3) now forces $\beta \neq 0$ so that we may assume without loss of generality that $\beta = 1$. We now have

$$D(a,b,c,d) = \frac{b + c}{\alpha a + b + c + \delta d}$$

where $\alpha > 0$ by (P4). We need only apply (P6) and (P0) to conclude that $\delta > 0$. ■

Definition. For any non-negative reals a and b, we denote by $D_{\alpha\delta}$ the PABDC identified by proposition 1:

$$D_{\alpha\delta}(a,b,c,d) = \frac{b+c}{\alpha a + b + c + \delta d} .$$

5. Global Order Equivalence and Monotonicity

Having classified all of the PABDCs satisfying (P0)-(P8), we now turn to the issue of distinctness. Naturally, for different α and δ values, the PABDCs determined in Proposition 1 would be distinct functions. But the actual dissimilarity *values* produced by a PABDC are often of less interest than the *rankings* of the dissimilarity values for a given object set. In fact, the entire class of Monotone Equivariant cluster methods is based on rankings rather than values produced by the dissimilarity coefficient involved (see Janowitz 1979). If we are concerned with rankings only, then we might be interested in finding alternative PABDCs which produce the same rankings but which exhibit particular mathematical properties such as metric or ultrametric properties (see Gower and Legendre 1986). We now define two relations between PABDCs that involve ranking equivalence.

Definition. Two PABDCs, D and D^* are said to form a *monotone pair* provided

$$\forall\, (a_1,b_1,c_1,d_1),\, (a_2,b_2,c_2,d_2) \in (\mathbb{Z}^+)^4 - \{(0,0,0,0)\}$$

with $a_1 + b_1 + c_1 + d_1 = a_2 + b_2 + c_2 + d_2$ we have:

$$D(a_2,b_2,c_2,d_2) < D(a_1,b_1,c_1,d_1) \text{ iff}$$
$$D^*(a_2,b_2,c_2,d_2) < D^*(a_1,b_1,c_1,d_1) .$$

Thus D and D^* form a monotone pair if they always produce the same rankings whenever the underlying attribute set is a given size. In particular, we can add new objects to the object set, and can change the recorded presence/absence for a particular object and attribute, and be assured that D will still produce the same ranking as D^*. In fact, we can even change the attribute set, as long as we do not change the number of attributes. If we are to require equivalent rankings even after changing the number of attributes, we need a stronger definition, taken from Sibson (1972) and rephrased for PABDCs:

Definition. Two PABDCs, D and D^*, are said to be *globally order equivalent* provided

$$\forall\,(a_1,b_1,c_1,d_1)\,,\,(a_2,b_2,c_2,d_2) \in (\mathbb{Z}^+)^4 - \{(0,0,0,0)\}\,,$$

we have

$$D(a_2,b_2,c_2,d_2) < D(a_1,b_1,c_1,d_1) \text{ iff}$$
$$D^*(a_2,b_2,c_2,d_2) < D^*(a_1,b_1,c_1,d_1)\,.$$

It should be noted that this definition agrees with global order equivalence in the sense of Janowitz (1979).

Proposition 2. *Given two PABDCs, $D_{\alpha\delta}$ and $D_{\alpha'\delta'}$, satisfying (P0)-(P8), the following statements are equivalent:*

(i) $\alpha / \delta = \alpha' / \delta'$;
(ii) *$D_{\alpha\delta}$ and $D_{\alpha'\delta'}$ are globally order equivalent;*
(iii) *$D_{\alpha\delta}$ and $D_{\alpha'\delta'}$ form a monotone pair.*

Proof. (i) $\Leftrightarrow$ (ii) Suppose $\dfrac{\alpha}{\delta} = \dfrac{\alpha'}{\delta'} = k$. Then

$$D_{\alpha\delta}(a_2,b_2,c_2,d_2) < D_{\alpha\delta}(a_1,b_1,c_1,d_1) \text{ iff}$$

$$\frac{(b_2+c_2)}{(b_2+c_2)+(\alpha a_2+\delta d_2)} < \frac{(b_1+c_1)}{(b_1+c_1)+(\alpha a_1+\delta d_1)} \text{ iff}$$

$$\frac{(b_2+c_2)}{(b_2+c_2)+\delta(k a_2+d_2)} < \frac{(b_1+c_1)}{(b_1+c_1)+\delta(k a_1+d_1)} \text{ iff}$$

$$(b_2+c_2)(ka_1+d_1) < (b_1+c_1)(ka_2+d_2)\,.$$

And, similarly

$$D_{\alpha'\delta'}(a_2,b_2,c_2,d_2) < D_{\alpha'\delta'}(a_1,b_1,c_1,d_1) \text{ iff}$$
$$(b_2 + c_2)(ka_1 + d_1) < (b_1 + c_1)(ka_2 + d_2)\,.$$

Hence $D_{\alpha\delta}$ and $D_{\alpha'\delta'}$ are globally order equivalent.

(ii) $\Rightarrow$ (iii) Trivial

(iii) $\Rightarrow$ (i) Let $k = \frac{\alpha}{\delta}$ and $k' = \frac{\alpha'}{\delta'}$. Note that $k,k' > 0$. Assume $k \neq k'$.

Without loss of generality, assume $k > k'$.

Select positive integers p and q such that $k > \frac{p}{q} > k'$.

Consider the four-tuples (a_1,b_1,c_1,d_1) and (a_2,b_2,c_2,d_2) where $a_1 = 0$, $b_1 = q + 1$, $c_1 = 0$, $d_1 = p(q + 1)$ and $a_2 = q(p + 1)$, $b_2 = p + 1$, $c_2 = 0$, $d_2 = 0$.

Then $\frac{b_2 d_1}{b_1 a_2} = \frac{p}{q}$, so that $k > \frac{b_2 d_1}{b_1 a_2} > k'$.

We therefore have $b_2 d_1 < k b_1 a_2$ and $b_2 d_1 > k' b_1 a_2$, which, together with our selection of (a_1,b_1,c_1,d_1) and (a_2,b_2,c_2,d_2) produce $(b_2 + c_2)(ka_1 + d_1) < (b_1 + c_1)(ka_2 + d_2)$ and $(b_2 + c_2)(k'a_1 + d_1) > (b_1 + c_1)(k'a_2 + d_2)$.

These inequalities are equivalent to $D_{\alpha\beta}(a_1,b_1,c_1,d_1) > D_{\alpha\beta}(a_2,b_2,c_2,d_2)$ and $D_{\alpha'\beta'}(a_1,b_1,c_1,d_1) < D_{\alpha'\beta'}(a_2,b_2,c_2,d_2)$ as was shown in the proof of (i) $\Rightarrow$ (ii).

Thus $D_{\alpha\beta}$ and $D_{\alpha'\beta'}$ do not form a monotone pair. This complete the proof. ∎

6. The Behavior of Some Common Dissimilarity Coefficients

Table 1 displays 21 fairly common similarity coefficients, the first 15 are taken from Gower and Legendre (1986). To examine behavior with respect to properties (P0)-(P8) we must first transform the coefficients into *dis*similarity coefficients in a reasonable way. That is, the ordering that each one induces on the domain must be inverted. Each conversion is done with one of several transformations as described below in such a way as to satisfy (P1) and (P3) if at all possible.

If the range of the similarity coefficient is [0,1] then we transform the coefficient by squaring it, if necessary, to resolve square roots, and then subtracting it from 1. In Table 1, coefficients 2-8, 10-13, 18, and 21 are transformed thus.

If the range of the similarity coefficient is [-1,1] then we transform the coefficient by subtracting it from 1 and dividing by 2. In Table 1, coefficients 9, 14, 15, and 17 are thus transformed.

Finally, coefficients 1 and 16 are transformed by adding 1 and inverting. Coefficients 19 and 20 are first multiplied by 4, then subtracted from 1.

Each of these dissimilarity coefficients has been checked against properties (P0)-(P8); Table 2 shows which properties are satisfied by these coefficients. In most cases, simple inspection suffices to determine whether a property is possessed by a DC, but in a few cases we rely upon the examination of a simple partial derivative of the DC in question, as indicated in the table.

Conclusions

Notice that the tables contain only 5 DCs which satisfy all 9 properties: DC#4 which, using the notation of section 4, is $D_{1,1}$, DC#6 which is $D_{0.5,0.5}$, DC#8 which is $D_{2,2}$, and DC#9 and DC#16 which are both $D_{1,1}$. All of these are globally order equivalent by Theorem 2. Note that the transformation of several distinct similarity coefficients produces identical dissimilarity coefficients. This happens simply because the coefficients in question are globally order equivalent, and we are concerned with rankings rather than actual numerical values when we perform the transformations.

One implication is that application of a Monotone Equivariant cluster method to data processed by any one of these four DCs will produce a clustering structure identical to that produced by applying the same cluster method to data processed by any other of these four DCs (see Janowitz 1979, Theorem 1). Hence, for computational and conceptual simplicity, we would recommend DC#4 over DC#6, 8, and 9 (provided that Monotone Equivariant cluster methods will be employed).

Evidently, it is beneficial to examine any heuristically constructed DC in the light of properties (P0)-(P8). Doing so could reveal undesirable behavior or indicate a simpler PABDC which is globally order equivalent to the original.

Finally, we would like to note that the axioms presented herein are intended for analysis of coefficients which are to apply to *all* data sets, in this way it is possible for one researcher to compare results in a meaningful way with the results of another researcher. We are not implying that those coefficients which do *not* satisfy these axioms are inferior to those which do,

No.	Similarity Coefficient	Dissimilarity Coefficient
(1)	$\frac{a}{b+c}$	$\frac{b+c}{a+b+c}$
(2)	$\frac{a}{a+b+c+d}$	$\frac{b+c+d}{a+b+c+d}$
(3)	$\frac{a}{a+b+c}$	$\frac{b+c}{a+b+c}$
(4)	$\frac{a+d}{a+b+c+d}$	$\frac{b+c}{a+b+c+d}$
(5)	$\frac{a}{a+2(b+c)}$	$\frac{2(b+c)}{a+2(b+c)}$
(6)	$\frac{a+d}{a+2(b+c)+d}$	$\frac{2(b+c)}{a+2(b+c)+d}$
(7)	$\frac{a}{a+\frac{1}{2}(b+c)}$	$\frac{b+c}{2a+b+c}$
(8)	$\frac{a+d}{a+\frac{1}{2}(b+c)+d}$	$\frac{b+c}{2a+b+c+2d}$
(9)	$\frac{a-(b+c)+d}{a+b+c+d}$	$\frac{b+c}{a+b+c+d}$
(10)	$\frac{1}{2}\left[\frac{a}{a+b}+\frac{a}{a+c}\right]$	$\frac{1}{2}\left[\frac{b}{a+b}+\frac{c}{a+c}\right]$
(11)	$\frac{1}{4}\left[\frac{a}{a+b}+\frac{a}{a+c}+\frac{d}{c+d}+\frac{d}{b+d}\right]$	$\frac{1}{4}\left[\frac{b}{a+b}+\frac{c}{a+c}+\frac{c}{c+d}+\frac{b}{b+d}\right]$
(12)	$\frac{a}{\sqrt{(a+b)(a+c)}}$	$\frac{(a+b)(a+c)-a^2}{(a+b)(a+c)}$

TABLE 1.

No.	Similarity Coefficient	Dissimilarity Coefficient
(13)	$\dfrac{ad}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}$	$1-\dfrac{a^2d^2}{(a+b)(a+c)(b+d)(c+d)}$
(14)	$\dfrac{ad-bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}$	$\dfrac{1}{2}\left[1+\dfrac{bc-ad}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}\right]$
(15)	$\dfrac{ad-bc}{ad+bc}$	$\dfrac{bc}{ad+bc}$
(16)	$\dfrac{a+d}{b+c}$	$\dfrac{b+c}{a+b+c+d}$
(17)	$\dfrac{\sqrt{ad}-\sqrt{bc}}{\sqrt{ad}+\sqrt{bc}}$	$\dfrac{\sqrt{bc}}{\sqrt{ad}+\sqrt{bc}}$
(18)	$\dfrac{(b-c)^2}{(a+b+c+d)^2}$	$\dfrac{n^2-(b-c)^2}{n^2}$
(19)	$\dfrac{bc}{(a+b+c+d)^2}$	$\dfrac{n^2-4bc}{n^2}$
(20)	$\dfrac{ad-bc}{(a+b+c+d)^2}$	$\dfrac{n^2-4(ad-bc)}{n^2}$
(21)	$\dfrac{(a+b+c+d)(b+c)-(b-c)^2}{(a+b+c+d)^2}$	$\dfrac{n^2-n(b+c)+(b-c)^2}{n^2}$

where we have let $n=(a+b+c+d)$ for conciseness in the last four DCs.

TABLE 1.

DC No. from table 1	(P0)	(P1)	(P2)	(P3)	(P4)	(P5)	(P6)	(P7)	(P8)
1	**N**	Y	Y	Y	Y	Y	Y	Y	Y
2	Y	Y	Y	Y	Y	Y	**N**	Y	Y
3	**N**	Y	Y	Y	Y	Y	Y	Y	Y
4	Y	Y	Y	Y	Y	Y	Y	Y	Y
5	**N**	Y	Y	Y	Y	Y	Y	Y	Y
6	Y	Y	Y	Y	Y	Y	Y	Y	Y
7	**N**	Y	Y	Y	Y	Y	Y	Y	Y
8	Y	Y	Y	Y	Y	Y	Y	Y	Y
9	Y	Y	Y	Y	Y	Y	Y	Y	Y
10	**N**	Y	Y	Y	Y	Y	Y	Y	**N**
11	**N**	Y	**U**	Y	Y	Y	Y	Y	**N**
12	**N**	Y	Y	Y	**Y***	Y	Y	Y	**N**
13	**N**	Y	**U**	Y	**N***	Y	**Y***	Y	**N**
14	**N**	Y	**U**	Y	Y	Y	Y	Y	**N**
15	**N**	Y	**U**	Y	**N***	**N****	Y	Y	**N**
16	Y	Y	Y	Y	Y	Y	Y	Y	Y
17	**N**	Y	**U**	Y	**N***	**N****	Y	Y	**N**
18	Y	Y	**N**	Y	**N**	**N**	**N**	Y	**N**
19	Y	Y	**N**	Y	**N**	**N**	**N**	Y	**N**
20	Y	Y	**N**	**Y****	**N**	**N**	**N**	Y	**N**
21	Y	Y	**N**	Y	**N**	**N**	**N**	Y	**N**

N= No, the DC does not satisfy this property (by inspection).
N*= No, the DC does not satisfy this property, since when d=0, a change in a has no effect.
N**= No, the DC does not satisfy this property, since when c=0, a change in b has no effect.
U= No, the DC does not satisfy this property, since its value is undefined at all points specified in (P2).
Y=Yes, the DC does satisfy this property (by inspection).
Y*= Yes, the DC does satisfy this property (easy to verify via partial derivative with respect to a.
Y**= Yes, provided a+b+c+d is even, No otherwise.

TABLE 2.

only that caution should be used when selecting them for use on a particular data set. Our axioms can be modified to suit other broad categories of data sets. For example: Suppose it is to be assumed that the attribute set, A, is augmented by a collection of "identity" attributes (for each object in the object set, add to A an artificial attribute which is "present" in that one object, and "absent" in all others). In such a situation, we would find that (P0) is too strong since under the new assumption, $d \neq n$, so that four-tuples such as (0,0,0,9) need not be in the domain of the coefficient. As a second example: If it is assumed that paired absences are to be ignored and that no objects will lack *all* attributes, then we can again weaken (P0) by not requiring that the coefficient be defined on such impossible tuples as $(0,0,0,d)$ where $d > 0$. We could also replace (P6) with an equality since paired absences are to be ignored.

A number of these special cases and variations on the set of axioms will be discussed in upcoming papers.

References

BAULIEU, F. B. (1985), "Order Theoretic Classification of Monotone Equivariant Cluster Methods," *Algebra Universalis, 20*, 351-367.

GOWER, J. C., and LEGENDRE, P. (1986), "Metric and Euclidean Properties of Dissimilarity Coefficients," *Journal of Classification, 3*, 5-48.

JANOWITZ, M. F. (1979), "Monotone Equivariant Cluster Analysis," *SIAM Journal on Applied Mathematics, 37*, 148-165.

JANOWITZ, M. F. (1979), "Preservation of Global Order Equivalence," *Journal of Mathematical Psychology, 20*, 78-88.

SIBSON, R. (1972), "Order Invariant Methods for Data Analysis," *Journal of the Royal Statistical Society, Series B, 34*, 311-349.