

ORDER-BASED STATISTICS AND MONOTONICITY: A FAMILY OF ORDINAL MEASURES OF ASSOCIATION*

LINTON C. FREEMAN

University of California, Irvine

This paper reviews standards for measures of association. In particular, standards for order-based measures are examined. The concept of monotonicity is shown to be ambiguous as it has been applied in this area, and it is clarified. The result is the specification of three — instead of the usual two — kinds of monotonic relations. These three monotonic models provide the basis for defining three non-arbitrary measures of ordinal association.

Eleven order-based measures are reviewed in the light of measurement standards as extended by the clarification of monotonicity. Eight are shown either to embody non-order-based elements or to contain *ad hoc* characteristics. The remaining three are shown to be members of a single family of monotone based ordinal measures that can be applied where ever their particular monotone models are appropriate.

INTRODUCTION

Often, perhaps too often, in the social sciences limitations in our theories or in our data put us in a position where we can specify only ordinal hypotheses. "The greater the A, the greater the B," we say, or "The greater the A, the less the B." The words, "greater" and "less" indicate that our observations must be at least ordinal. But the lack of

* I want to express my deep appreciation to those people who gave me enough significant feedback on earlier versions of this paper to prevent me from rushing into print with more, and more important, errors than the present version still contains. Among these, Katie Faust, Sue Freeman, Timlynn Babitsky, Mike Migalski and Tom Nelson are outstanding.

specificity in the hypotheses themselves suggest that — even if we have recorded our data at an interval or ratio level — it would be inappropriate to use a measure of association based on a linear equation or any other stronger than ordinal model. Indeed, data cannot be brought to bear on such *inherently ordinal hypotheses* unless we have appropriate statistical tools — tools that permit the direct examination of the order in relationships. So in these cases we need statistics that are responsive to order but that do not depend on the magnitudes of the variable we observe.

Fortunately for us, various statisticians and social science methodologists have put in a good deal of effort toward developing strictly order-based measures of association (see Kendall, 1948; Kruskal, 1958; and Goodman and Kruskal, 1959 for reviews of much of this literature).

Anyone who has ever called for all the statistics in the Frequency Tables Program of BMDP or the CROSSTABS Program of SPSS knows that there are a great many seemingly order-based measures that are available for use. The only trouble is that with so many it is difficult to make a reasonable choice among them in a particular application. Moreover, as we shall see, much of the literature surrounding these statistics is confused and confusing. Many turn out to be not strictly order-based. Currently, then, choosing among ordinal measures on a rational basis is just about impossible.

The present paper is intended to contribute to clarifying ambiguities about the measurement of ordinal association. It will draw on and extend ideas introduced by Goodman and Kruskal (1954) and Costner (1965). Here the emphasis will be on problems of modelling ordinal association and, in particular, on monotonic models. Some older order-based statistics will be re-examined. And finally, a family of three acceptable ordinal measures will be specified.

In the first section a set of established standards for measuring ordinal association will be reviewed. That review will be followed by a clarification of the concept of monotonicity. Then a number of standard measures will be examined in the light of both established standards and the clarified monotonicity idea. Finally, three measures will be shown to be members of an interrelated family of non-arbitrary indices of ordinal association.

STANDARDS: WHAT ON ORDINAL MEASURE SHOULD LOOK LIKE

Explicit standards for measuring association have been developed and set down in papers by Goodman and Kruskal (1954) and Costner (1965). The Costner treatment, however is more general. It is intended to set standards for *any* measure of association, while Goodman and Kruskal limited their focus to measuring association on observations recorded in cross-classification tables. It is appropriate therefore to begin with more general Costner treatment.

Costner (1965) developed an earlier idea by Guttman (1941) and proposed that any measure of association — ordinal or otherwise — should exemplify a general proportional reduction in error (PRE) model of the following form:

$$\text{PRE} = \frac{E_1 - E_2}{E_1}$$

where E_2 is an index of error in guessing some property of a variable in the light of information about another variable, and E_1 is the same index used when guesses are made in the absence of such information. A PRE index, then, is an error ratio. It takes a value of 0 only when information yielded by the predictor variable is of no help. It is 1 only when information from the predictor variable eliminates all error. And it can always be interpreted as the proportion of error in making guesses about a variable that can be eliminated by taking information provided by another variable into account.

Today the PRE notion is widely accepted. Observers agree that the PRE idea results in a very powerful convention and that it can be used as a basis for evaluating any measure of association. It guarantees a meaningful index of the *amount* or *degree* of association. If, for example, a PRE measure yields a result of .57, we can answer questions about what that number means. It means precisely that 57% of our error in guessing something about Y can be eliminated when we take something about X into account.

But the PRE convention, though powerful, still leaves something out. It provides no guidance on the interrelated problems of what it is about a variable that we are to guess and how we should measure error. The solutions to these problems rest in specifying explicit models for "perfect association" and for scaling departures from that model in a meaningful way. When we have both an association model and an error model, the rules for guessing and for measuring error should follow.

Without such a model, the door is open for the construction of completely arbitrary and *ad hoc*, though well scaled, measures.

Costner was aware of at least part of this problem. He said that "the meaning of a relationship of a specified 'degree' remains ambiguous unless, implicitly or explicitly, its 'form' or 'shape' is specified." When it came to dealing with measures, however, his specifications were always implicit. Nowhere in the Costner paper is there an explicit discussion of the form or shape of the relationship on which any order-based measure is based.

The paper by Goodman and Kruskal (1954) comes closer to addressing this issue, at least so far as constructing ordinal measures is concerned. They proposed that an ordinal measure should be based on the probabilistic notion of the 'optimal prediction of order.' The word "optimal" here refers to some notion of perfect association, but, again, their model was implicit. In their own words, "There is vagueness in the idea of completely ordered association."

Nevertheless, Goodman and Kruskal did present a systematic way of looking at relationships between two ordered polytomies. Let us assume, they suggested, that we have data that can be displayed in an ordered contingency table containing cross tabulations of observations on two variables, X and Y. Both X and Y are measured at, at least, the ordinal level. Thus, $X = \{x_1, x_2, \dots, x_N\}$ where $x_i > x_j$, $x_i < x_j$ or $x_i = x_j$ for all x in X. $Y = \{y_1, y_2, \dots, Y_N\}$ must, of course, be similarly ordered.

Any ordinal measure must, from this perspective, focus somehow on the relationship between order in X and order in Y. This cannot be done without *comparing* observations in terms of which is greater and which is less on each variable. In particular, it is convenient to focus on *pairs* of observations to determine the degree to which they are in the same or in the inverse order on the two variables. To develop the Goodman and Kruskal perspective (and for later discussion of particular measures) it will be useful to distinguish among various kinds of pairs that might be found in ordered contingency tables. In particular, we will need the following ideas. Let

N = number of cases observed

and

P = total number of pairs of observations
= $(N^2 - N)/2$.

Now P may be partitioned into five kinds of pairs:

- C = number of concordant (same order) pairs
 D = number of discordant (inverse order) pairs
 X_0 = number of pairs tied only on X
 Y_0 = number of pairs tied only on Y
 Z = number of pairs tied on both X and Y

and therefore,

$$P = C + D + X_0 + Y_0 + Z.$$

Goodman and Kruskal presented an intuitive conception of optimal order. They argued — at least indirectly — that if all pairs in a contingency table are tabulated in C — if every pair falls such that $x_i > x_j$ and $y_i > y_j$, there is some kind of “perfect association” between X and Y. † Similarly, if all pairs are tabulated in D — if for every i and j, $x_i > x_j$ and $y_i < y_j$ then again “perfect association” obtains. They proposed then, that these two “optimal” orders could be used as the basis for constructing measures.

In cases where neither of these pure orders is exhibited, we need a way of determining the degree to which one or the other dominates. We can choose pairs of observations at random (with replacement) and record whether each pair contributes to the C tally or to D. Those associated with C contribute toward one ideal type and those associated with D to the other. So an index based on the *difference* between C and D reveals the degree to which the observed data approach one or the other ideal. Constructing a measure, then is a straightforward task of balancing C off against D and seeing how much the arrangement of the data approaches one pure order or the other.

A raw index like $C - D$ is obviously inappropriate. But as an alternative, Goodman and Kruskal proposed that two probabilities be calculated, the probability of C and the probability of D. Their model for an ordinal measure of association, then, is simply the difference between these two probabilities, $\Pr(C) - \Pr(D)$.

Clearly, any measure consistent with this model is order-based. Its magnitude depends *only* on the degree to which the observations approach one or the other ideal form of ordered relationship. Under statistical independence, since $C = D$, the measure yields a value of 0. It

† Throughout this paper I will refer to pairs of observations on X and Y. When orders are compared, I will specify $x_i > x_j$ and $y_i > y_j$ to describe concordance or agreement in order. Strictly speaking, one might add the case where $x_i < x_j$ and $y_i < y_j$ as the “other half” of concordance. But, since we are dealing with unordered pairs, we can always exchange the labels, i and j, so the first expression fits and the “other half” is unnecessary as well as redundant.

grows only as departures from independence are ordered — one way or the other. It grows, that is, only with *ordinal* departures from independence, as C and D increasingly depart from the equality. When C is at its maximum and $D = 0$, all pairs fall in the same order in X and Y and the measure takes a value of 1. And when D is at its maximum and $C = 0$, all pairs fall in inverse order and the measure is -1 . The magnitudes of values falling between 0 and an absolute value of 1 indicate how much more likely it is to see the dominant as opposed to the non-dominant order between the two variables.

Thus, the Goodman-Kruskal approach to ordinal measurement suggests the direction in which to seek an answer to the problem of specifying the "shape" of an order-based measure of association. But this approach still leaves a certain degree of ambiguity. Although it provides a model for taking order — and order alone — into account when constructing a measure, it does not provide a general solution to the problem of what is meant by completely ordered association.

The problem is that there are several ways to calculate the probabilities of concordant and discordant pairs. Goodman and Kruskal suggest that they be calculated conditional on both x_i not $= x_j$ and y_i not $= y_j$ — that all pairs of observations that are tied in either X or Y (or both) be excluded from computations. Other writers that accept the general Goodman and Kruskal approach have proposed that only pairs that are tied in X be excluded, or only those that are tied in Y or whatever. Clearly, each of these conventions for calculating $Pr(C)$ and $Pr(D)$ embodies a different model of completely ordered association. What is still lacking is a general way of thinking about alternative notions of the shape or form of complete association in the ordinal case.

More recent attempts to confront this problem have drawn on the mathematical idea of monotones (Somers, 1962; Leik and Gove 1969, 1971). But monotonicity, it turns out, is not well understood by these writers. The various notions of monotonicity contained in this literature are for the most part somewhat loose and incomplete. As a consequence, no systematic overall model or set of models for ordinal association have been developed. Some apparently order-based measures are not truly based on order, while others have a distinctly *ad hoc* character.

In the next section, therefore, we shall examine the idea of monotonicity and attempt to develop a systematic basis for thinking about the form or shape of complete ordinal association.

ON MONOTONICITY

Mathematically, monotones are kinds of functions. There are two broad classes of monotonic functions, strong (or strict) and weak. Then each of these classes is partitioned into two more subclasses according to direction. Thus, among strong monotonic functions, one is called increasing and one decreasing. Correspondingly, one of the weak monotonic functions is called non-decreasing and the other non-increasing.

An increasing monotonic function is a subset of the X by Y Cartesian product such that for every pair of elements from X , where $x_i > x_j$, then $y_i > y_j$. If the function is decreasing, given a pair of elements in X where $x_i > x_j$, then $y_i < y_j$. And, of course, if these functions are *onto*†, if $x_i = x_j$, then $y_i = y_j$ and both increasing and decreasing monotones are one-to-one correspondences. Furthermore, a function is non-decreasing if for every pair of elements in X , where $x_i > x_j$, then y_i not $< y_j$, and non-increasing if, for every $x_i > x_j$, y_i not $> y_j$.

Intuitively, all this permits the description of various kinds of ordered relationships. Increasing and non-decreasing monotonic functions are direct; X and Y grow together. Decreasing and non-increasing ones are inverse; they grow in opposite directions. Strong monotonic functions are characterized by regular growth (or, if the function is decreasing, decay); thus, increases in X are always accompanied by increases (or, for decreasing functions, decreases) in Y and vice-versa. Weak monotonic functions, in contrast, capture a situation of possible step-wise growth. Increases in X are accompanied by either increases (or, for the non-increasing function, decreases) or by a lack of change in Y .

Increasing and non-decreasing monotonic functions are illustrated in Figure 1. As the pictures show, since they are functions, all of these mathematical monotones are at least many-to-one relations. Thus they are mappings where there is a unique element in the range for every element in the domain.

This same vocabulary is used by statistical writers in discussing order-based measures. The only problem is that the words are used in a slightly different way. They would all agree, I suspect, that what the

† In the statistical analysis of contingency tables, such functions must always be *onto* since otherwise our tables could have an indeterminate number of blank rows and/or columns. Such rows or columns that contain no non-zero entries can contribute nothing to the analysis of ordinal association. So if we simply drop empty rows or columns before analysis begins we will always be dealing with onto mappings.

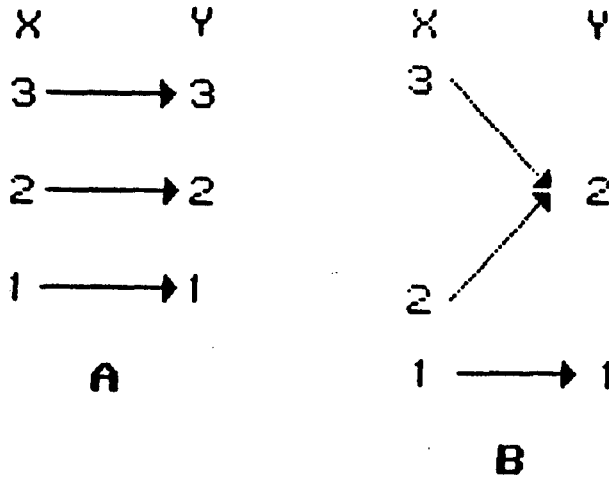


FIGURE 1. A — Strong Monotonic Function
B — Weak Monotonic Function

mathematician calls a strong monotone is a strong monotone. But the mathematician's weak monotone seems also to be classed as strong (Leik and Gove, 1969) and the expression "weak monotone" is sometimes reserved for an entirely different object.

At first glance, the statistician's weak monotone looks exactly like the one defined above. Given X and Y as above, we have a statistician's non-decreasing relation if for every pair of elements in X , $x_i > x_j$, $y_i \geq y_j$ and a non-increasing one if for every $x_i > x_j$, $y_i \leq y_j$. But there is a critical difference. While the mathematician restricts the definition to functions, the statistician includes non-functional or many-to-many relations.

Figure 2 shows a mapping that forms the monotonic basis for Goodman and Kruskal's (1959) statistic, gamma. It is monotonic in the important sense that observations in X and Y are in the same order for all pairs in which both X and Y observations are ordered. There are no inversions in order. But neither the mapping shown in Figure 2 nor its inverse is a function. It is monotonic, but not in the usual mathematical sense.

This suggests that from the perspective of providing models of ordinal relationships we have not two, but three, basic kinds of monotones (and, of course, two directions for each). The mathematician's strong monotone is a *monotonic one-to-one correspondence* from this perspective. The mathematician's weak monotone is a

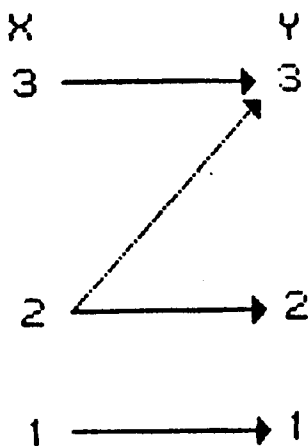


FIGURE 2. Statisticians' Monotone

monotonic function. And the statistician's monotone is a *monotonic relation*.

The implications of these ideas for data analysis can be seen by translating them into contingency tables. Figure 3 shows tables that embody the three kinds of monotonicity described here. An entry of "P" in a cell indicates a permissible observation according to the relevant conception of monotonicity, and cells are left blank where observations that violate the model might fall.

By inspecting these tables we can see, in each case, how various kinds of ties must be treated under the conditions imposed by that monotonic model. In all three cases observations that are tabulated in Z are permissible. Those are the pairs that are contained within a single cell of a table. In effect, they are nothing but "replications." As such, they neither contribute to nor detract from the degree to which the data fit the model. The problem, then, is how to deal with X_0 and Y_0 ties.

X_0 ties are generated by pairs that fall in different rows but the same column of the table. In such cases, pairs are tied on the X variable, but ordered on Y. Correspondingly, Y_0 pairs fall in different columns but the same row and are tied in Y. The model of monotonic one-to-one correspondence requires that neither of these kinds of ties be present. If either kind appears, the relation is not one-to-one. The monotonic function, however, when mapping from X to Y, excludes only pairs counted in X_0 . Such pairs that are tied in X indicate a mapping from one point in X to two in Y. Such a mapping is not many-to-one. However,

pairs that are tied in Y are not harmful under this model. They are part of the "many" in the many-to-one mapping and as such are permissible. In the monotonic relation model both kinds of ties are permissible. Since this is a many-to-many mapping, neither kind of tying is prohibited under the model.

If we combine these conceptions of monotonic relations with the Goodman and Kruskal probabilistic model of an order-based measure, we can specify exactly the forms that monotonic order-based measures must take. We end up, then, with three models of complete ordinal association. The one based on a monotonic relation must have the form

$$\frac{C}{C + D} - \frac{D}{C + D}.$$

The monotonic function from X to Y must have the form

$$\frac{C}{C + D + X_0} - \frac{D}{C + D + X_0}.$$

And the monotonic one-to-one correspondence must have the form

$$\frac{C}{C + D + X_0 + Y_0} - \frac{D}{C + D + X_0 + Y_0}.$$

These three kinds of monotonic order-based models exhaust all the possibilities. If we remember that we could calculate the monotonic function either on a data matrix or its transpose (thereby using Y_0 rather than X_0 in the denominator), it is clear that no other kinds of measures that are strictly order-based could exist. Indeed, when PRE realizations of each of these models is specified, the job is done. Other strict ordinal measures that meet the criteria specified here cannot exist.

These three models embody explicit specifications of what we mean by a perfect relationship. Thus, the problem of specifying the form or shape of an ordinal relationship is solved without asserting *ad hoc* rules for handling tied observations. Given these models, we know how to treat ties and we can generate measures in which we can make explicit not only our model for degree of association, but our model for its form as well. All that remains to be done is to find or construct measures that comply with these models and show that they comply also with Costner's PRE standard. In the next section, therefore, we will review some existing measures and see how they "stack up" in the light of these new criteria.

ORDINAL MEASURES: A REVIEW

In an exhaustive review of the older literature, Kruskal (1959) showed that the idea of measuring ordinal association began around the turn of the century with work by Galton. But the earliest measure that still is in current use is the *rho* that was developed by Spearman (1904). Essentially, *rho* was introduced as a convenient computing formula for Pearson's *r* when the data were made up of the first *N* natural numbers. Kruskal (1958) has shown that *rho* can be given an order-based PRE interpretation, but that interpretation involves counting orders, not only on pairs, but on 3-tuples as well.

Costner (1965) saw a problem in this kind of counting. He rejected *rho*, arguing that it "may be interpreted in a proportional-reduction-of-error sense only by specifying rather complex rules and definitions, and such complexity suggests that only in very few, if any, actual research situations would the investigator be interested in the relative accuracy of estimation." But Costner saw only part of the problem. The kind of counting involved in *rho* results in a completely arbitrary weighting of concordances and discordances. The result is a statistic that can depart from 0 even when *C* is exactly equal to *D* — when there is no ordinal departure from independence. Thus, *rho* is not simply order-based.

Kendall (1938) introduced an alternative to *rho* called *tau*. *Tau* was implicitly based in the idea of a monotonic one-to-one correspondence. It was defined by Kendall as

$$\tau = \frac{C}{P} - \frac{D}{P} = \frac{C - D}{P}$$

where all observations on both variables are strictly ordered; where either $x_i > x_j$ or $x_i < x_j$ for all x in X and similarly for all y in Y . Thus, given these conditions, $X_0 = Y_0 = Z = 0$, P is simply the sum of C and D and *tau* is an index of the overall tendency of pairs to be in the same or the inverse order in X and Y .

From the PRE perspective we think of *tau* as involving two kinds of guesses. First we guess the order of pairs of observations in the dependent variable (Y) without reference to the independent variable (X). So we randomly choose pairs of observations from the Y distribution. We shall guess that $y_i > y_j$ with a probability of 1/2. Our expected error from this operation is $(C + D)/2$. This is E_1 . Now we repeat this guessing operation with X taken into account. Here we look at C and D , and if $C > D$ we guess concordance uniformly for all pairs; otherwise

we guess discordance. The expected error, E_2 , is either C or D , whichever is smaller. Assume $C > D$. Then

$$\begin{aligned} \text{PRE} &= \frac{(C + D)/2 - D}{(C + D)/2} \\ &= \frac{C - D}{C + D}. \end{aligned}$$

Since without ties, $X_0 = Y_0 = Z = 0$, then $P = C + D$, and we end up with tau. Of course, if $D > C$, the numerator becomes $D - C$, but in general, if we let it be $C - D$, the PRE version can take the same sign as tau as defined by Kendall.

All in all, then, tau seems to meet all our criteria for an ordinal measure of association. Its application, however, is restricted to those situations where our data show no ties in either X or Y . This is a serious limitation and it means that tau is not the sort of general ordinal statistic we are seeking.

So the search began. First Kendall himself, then a whole host of other statisticians and research scientists set out on a quest for generalizations of tau that would relax the restriction on tied observations and still permit the systematic study of order relations between variables.

The measure called τ_b was Kendall's (1948) own attempt to grapple with this problem. The measure

$$\tau_b = \frac{C - D}{\sqrt{(C + D + X_0)(C + D + Y_0)}}$$

has as its denominator the geometric mean of the number of pairs in the marginal distribution of Y that are not tied on Y and the similar number for X . That denominator was designed to permit τ_b to take a value of 1 in cases of "perfect association." It was not at that time at all clear what "perfect association" meant, so the resulting calculations were from the current perspective simply *ad hoc*. To make matters worse, this denominator failed in its attempt to give the statistic an upper limit of 1 in all cases.

τ_c was suggested by Stuart (1953) in order to compensate for a perceived "weakness" of τ_b , the fact that it could never reach a value of 1 when the number of rows in the table on which it was computed was unequal to the number of columns. Stuart "corrected" for this "problem" by constructing a new denominator that was a function of the minimum of the number of rows and the number of columns. This correction permits τ_c to achieve a value of "almost" 1. But Stuart's efforts were completely misguided. The whole point is that he and

Kendall were seeking a monotonic one-to-one correspondence, and such a monotone can exist *only* when the number of rows is precisely equal to the number of columns. Indeed, Stuart's statistic is completely without a monotonic relationship model of any sort.

Goodman and Kruskal's (1954) *gamma* is a straightforward order-based measure of association based on the model of the monotonic relation. From the discussion above we know that a monotonic relation model must be of the form

$$\frac{C}{C + D} - \frac{D}{C + D} = \frac{C - D}{C + D},$$

since X_0 , Y_0 and Z are all excluded. And that is exactly the expression that Goodman and Kruskal used to define gamma. This means that, in gamma, ordinal predictions are made only for those pairs of observations that are untied. Pairs are chosen only if they are not tied either in X or in Y and gamma is the difference between the conditional probabilities of like and unlike orders under that choice rule.

When X and Y are statistically independent, $C = D$ and gamma is 0. Gamma grows, not with just any departures from independence, but only with *monotonic* departures — one way or the other. As the preponderance of pairs of untied observations tend systematically to be in the same or in the inverse order on X and Y , gamma departs from 0.

The sign of gamma, plus or minus, indicates whether the dominant order is the same or the inverse on the two variables. And its magnitude can be directly interpreted as telling us how much more likely we are to see same rather than inverse order in untied observations on the two variables. Thus, a gamma of +1 indicates that all untied pairs are in the same order, and -1, that they are all in inverse order. As a final bonus, gamma, or rather its numerator, has a known and tractable sampling distribution.

Gamma is also a PRE measure and is derived by exactly the same reasoning as we used above in providing a PRE basis for tau:

$$\begin{aligned} \text{Gamma} &= \frac{(C + D)/2 - D}{(C + D)/2} \\ &= \frac{C - D}{C + D}. \end{aligned}$$

But here, ties are not prohibited — they are simply not used in calculations. So gamma does not require a one-to-one correspondence, or even a function, to achieve a value of 1. The magnitude of gamma will be 1 whenever the observations can be fit to any monotonic relation.

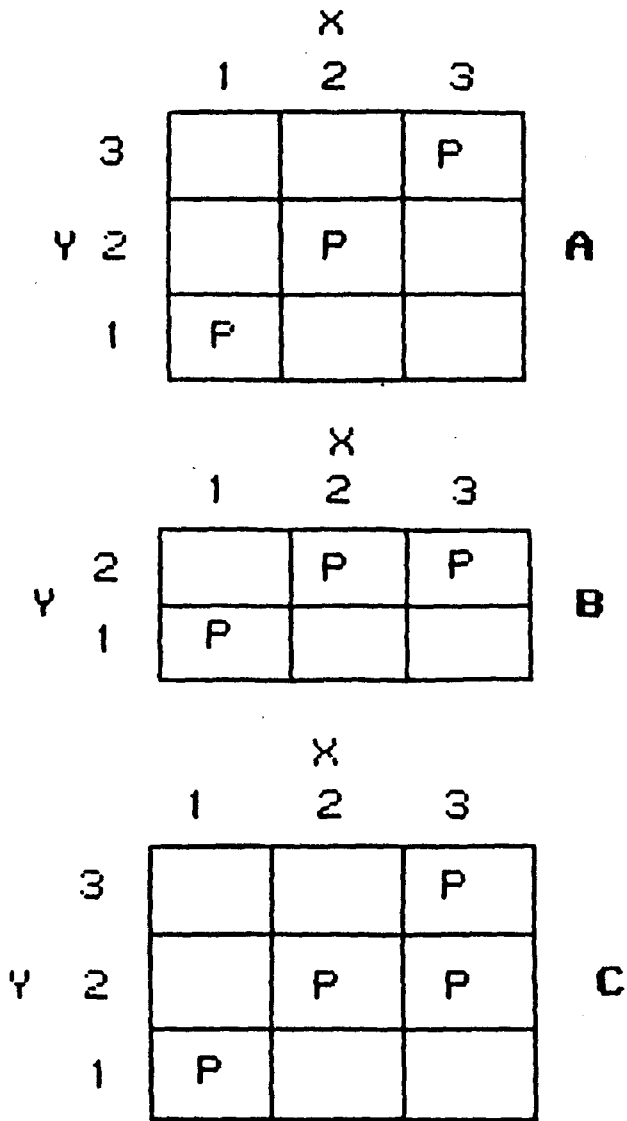


FIGURE 3. A — Monotonic one-to-one
 B — Monotonic function
 C — Monotonic relation

Gamma is, therefore, a perfectly general and interpretable measure of ordinal association that can be used whenever the monotonic relation is an appropriate model for the problem and data at hand.

Some writers have criticized gamma for the "undesirable property" of restricting its calculations only to untied pairs (Loether and McTavish, 1976). This property, however is neither a flaw nor a weakness. It is an unalterable consequence of gamma's association model. Indeed, it might be more appropriate to complain that, in a given application, the monotonic relation is not the model desired.

With gamma, then, we have our first unrestricted and completely non-arbitrary measure of ordinal association. It is based on the model of a monotonic relation and it is both strictly order-based and a PRE measure.

The next several measures to be reviewed are all asymmetrical. All, therefore, are — at least implicitly — aimed at embodying the monotonic function as their model.

Sommers (1962) explicitly sought to create a measure based on the notion of monotonicity. He introduced an asymmetric measure, d_{yx} , but did not specify just how it tied in to the monotone notion. d_{yx} was defined as an asymmetrical measure where X is the independent and Y the dependent variable. It is equal to

$$d_{yx} = \frac{C}{C + D + Y_0} - \frac{D}{C + D + Y_0} = \frac{C - D}{C + D + Y_0}.$$

On the face of it, the Sommers measure seems to be the opposite of the measure we are seeking. The model of a monotonic function excludes Y_0 pairs, but Sommers excludes those that are tallied in X_0 . His reasoning seems similar to that underlying τ_b . The denominator of d_{yx} is equal to the number of pairs that are not tied in the marginal distribution of X — the number that might possibly contribute to C or D. So, from this perspective, d_{yx} is a measure of the degree to which the predominant mode, same or inverse order, approaches its maximum value. Or, from the viewpoint of Goodman and Kruskal, it is the difference between the probabilities of like and unlike orders conditional on pairs being chosen that are not tied on the independent variable, X. But from the perspective of modelling a monotone, the wrong ties are excluded. As defined, then, d_{yx} does not fit the model of a monotonic function.

Moreover, we can see that as a consequence of its not being based on a monotonic model d_{yx} is seriously flawed. Leik and Gove (1969) have shown that although d_{yx} can be given a PRE interpretation, calculating

errors depends on an awkward weighting scheme. Half the pairs that are tied on the dependent variable — Y — but not tied on X , are defined as errors. Thus, if two observations in X are ordered, a tie between the corresponding pair in Y is considered to be half as much of an error as a clear reversal in order in the Y pair. This kind of arbitrary differential weighting of various kinds of errors gives the resulting measure a distinctly *ad hoc* character. It defeats the whole aim of developing systematic interpretable measures.

Since Leik and Gove (1969) were troubled by the arbitrary weighting in the Sommers statistic, they introduced a variant, $d'_{y,x}$. There, errors generated by Y_0 were counted in a straightforward way rather than being weighted by $1/2$. This measure was said to be designed as a "strong monotone" version of $d_{y,x}$, but, like $d_{y,x}$, it was also reversed with respect to fitting a monotonic function. Moreover, as a consequence of their rule for counting errors, Leik and Gove produced a measure that cannot be interpreted as a difference in the conditional probabilities of like and inverse orders; theirs is not a strictly ordinal measure. Thus, $d'_{y,x}$ represents an attempt to repair $d_{y,x}$ that was focussed on the wrong problem.

Kim (1971), however, found the right problem. His $d_{y,x}$ is an asymmetrical measure from X to Y . It requires that we choose pairs conditional on y_i not = y_j . This eliminates pairs that fall in Y_0 and Z from consideration. The initial guessing rule is the same as that described above. Guess orders in the dependent variable, $y_i > y_j$ with a probability of $1/2$. Since all $y_i = y_j$ pairs have been excluded, our expected error is $1/2$ for all the remaining pairs or $1/2 (C + D + X_0)$. This is E_1 .

The standard rule is used for guessing orders of pairs of Y observations, given information on X pairs. If $C > D$ and $x_i > x_j$, guess $y_i > y_j$. Here the pairs that are tied in X are included, so guesses have to be made for them. But since for these pairs $x_i = x_j$, knowledge of the X pair gives us no new information; we can still do no better than to flip a fair coin and guess $y_i > y_j$ or $y_i < y_j$ with a probability of $1/2$. Thus, our second error index, $E_2 = D + X_0/2$. Then

$$\begin{aligned} d_{y,x} &= \frac{(C + D + X_0)/2 - (D + X_0/2)}{(C + D + X_0)/2} \\ &= \frac{C - D}{C + D + X_0} \end{aligned}$$

Thus, $d_{y,x}$ is both a PRE measure and is strictly order-based from the Goodman-Kruskal perspective. It involves no arbitrary weighting, and

it embodies the monotonic function as its ordinal model. All this means that the Kim asymmetrical statistic may be used whenever the idea of a monotonic function is appropriate to the problem at hand. With Kim's $d_{y,x}$ we have, therefore, our second general and non-arbitrary measure of ordinal association.

Before leaving this discussion of asymmetrical measures, we must mention one other case, Leik and Gove's (1971) index, d_0 . The creation of d_0 was motivated by Costner's argument that, because C, D and possibly X_0 and Y_0 were used in calculating his first error index, E_1 , these models were not strictly PRE. Orders on the predictor variable, X, seemed to be used in determining orders on the predicted variable, Y, while in the E_1 case, guessing is supposed to involve only orders on Y. Actually, the orders of X are only used for computing errors — they are not used for making guesses so the apparent problem is not really substantive. Nonetheless, Leik and Gove set out to create a measure that eliminated that "problem."

The result of this effort is a measure that is certainly PRE and that seems, at first glance, to resemble a Goodman-Kruskal order-based measure. The trouble is that it is actually based, not on optimal orders, but rather on departures from statistical independence. That being the case, d_0 — like rho — can be shown to depart from 0 in cases where there is no monotonic departure from independence — where $C = D$. d_0 , then is not a strictly order-based statistic. It contains elements that make it impossible to interpret in the context of ordinal measurement.

We are left at this point still looking for the statistic Kendall sought — an order-based measure based on the monotonic one-to-one correspondence model. Kim (1971) proposed a symmetrical version of his $d_{y,x}$ that is a possible candidate. His symmetrical measure, d , is simply the average of two asymmetrical measures, $d_{y,x}$ and $d_{x,y}$. Thus, it is easy to show that

$$d = \frac{C}{C + D + X_0/2 + Y_0/2} - \frac{D}{C + D + X_0/2 + Y_0/2}$$

$$= \frac{C + D}{C + D + X_0/2 + Y_0/2}$$

The result is an arbitrary weighting of errors generated by X_0 and Y_0 and a measure that does not capture the monotonic model of interest.

Finally, we turn to a symmetrical measure, e , introduced by Wilson (1974) and according to Kruskal (1958) also by Deuchler (1914). Like the others, this measure is based on the order exhibited by pairs of observations. And like the others, this order is guessed for each pair of

observations for a variable both with and without information about the other variable. The difference here is that orders are guessed twice: first for Y and then for X. The orders of Y pairs are guessed both alone and taking X orders into account, and the orders of X pairs are similarly guessed both alone and taking Y orders into account.

In guessing Y orders, we choose pairs of observations conditional on x_i not $= x_j$; that is we exclude pairs that are tallied in X_0 and Z. This is necessary in order to avoid counting those pairs as errors. We will pick up the pairs in X_0 when we guess the orders in X, and we do not want to count the Z pairs as errors at all. As always, we guess $y_i < y_j$ with a probability of one-half. We will make errors in such predictions of Y for 1/2 of those pairs counted as C and 1/2 of those counted as D. In addition, all cases where $y_i = y_j$ — those tallied in Y_0 — are errors. Therefore, for the Y predictions,

$$E_1(Y) = (C + D)/2 + Y_0.$$

Exactly the same procedure may be used to guess orders in X without knowledge of Y. For these X predictions, then,

$$E_1(X) = (C + D)/2 + X_0,$$

and the total error in guessing orders on each variable without knowledge of the other is

$$\begin{aligned} E_1 &= E_1(Y) + E_1(X) \\ &= C + D + X_0 + Y_0. \end{aligned}$$

In predicting the order in Y conditional on the X order, we must again exclude those pairs that are tied on X. If $C > D$, we guess

$$y_i > y_j | x_i > x_j$$

that is, we guess concordance uniformly for all pairs. Using this rule we make errors for all pairs that are tallied in D and all those where we make ordinal predictions but observe ties — those tallied in Y_0 . Thus, $E_2(Y) = D + Y_0$.

Now we guess the order of X based on the Y order by the corresponding rules, and errors occur for pairs tallied in D and those in X_0 . Error for this set of predictions, then, is $E_2(X) = D + X_0$, and the total error in both variables taking the other into account is

$$E_2 = 2D + X_0 + Y_0.$$

Therefore, as a general PRE measure,

$$\begin{aligned}
 e &= \frac{E_1 - E_2}{E_1} \\
 &= \frac{(C + D + X_0 + Y_0) - (2D + X_0 + Y_0)}{C + D + X_0 + Y_0} \\
 &= \frac{C - D}{C + D + X_0 + Y_0},
 \end{aligned}$$

which is exactly the measure we seek.

Thus e is both a PRE measure and a strictly order-based measure. It takes as its model the general monotonic one-to-one correspondence. It is therefore a straightforward generalization of Kendall's tau. It lacks the restriction that $X_0 = Y_0 = Z = 0$ that characterizes tau, but for data where that restriction holds tau is equal to e . Wilson's e statistic may be used to determine the degree to which a monotonic one-to-one correspondence fits our data whenever we have a bivariate distribution and our observations are recorded at least at the ordinal level.

This ends our review of standard measures of ordinal association. We end up with three broadly useful measures, gamma, $d_{y,x}$ and e . At the level of models of association, gamma embodies a monotonic relation, $d_{y,x}$ a monotonic function and e a monotonic one-to-one correspondence. They are a family in the sense that they all exhibit the same form. All, as a matter of fact, have exactly the same numerator. All therefore are strictly order based; they are differences between conditional probabilities of like and inverse orders.

These three measures differ only in their denominators, where each "penalizes" the data for tied observations in its own way and consistent with its own monotone model. And that's where the conditions on the probabilities come in. Gamma is conditional on x_i not = x_j and y_i not = y_j . $d_{y,x}$ is conditional on y_i not = y_j . And e is conditional on x_i not = x_j when guessing Y and y_i not = y_j when guessing X .

SUMMARY AND CONCLUSION

This paper began with a review of an explicit and generally accepted set of standards for measuring ordinal association. These standards, however, were shown to contain pockets of ambiguity that could result in the construction of measures that complied totally with their rules but that still contained arbitrary and *ad hoc* elements.

It was argued that what was needed was a standard or standards for the notion of completely ordered association. And the answer was sought in the mathematical concept of monotonicity. Three distin-

guishable kinds of monotones — a monotonic relation, a monotonic function and a monotonic one-to-one correspondence were developed. These three monotones were cast as models of complete association that could form the bases for three measures of ordinal association. This paper, then, extended the current standards for measuring ordinal association in order to eliminate their residual ambiguity.

A review of eleven established measures showed eight that failed to meet these revised standards and three that succeeded. The successful measures were Goodman and Kruskal's (1954) gamma, Kim's (1971) $d_{y,x}$ and Wilson's (1974) e . All the others contained *ad hoc* elements or other limitations that preclude their use as general ordinal measures.

The result is a family of interpretable monotone based measures. They are all related — in fact, they differ only in their denominators. It is always the case that $e \leq d_{y,x} \leq \text{gamma}$. Any of these statistics may be used to examine data in applications where its monotonic model is consistent with the substantive hypotheses.

REFERENCES

- Costner, H.L. 1965. Criteria for measures of association. *American Sociological Review* 30, 341-353.
- Deuchler, G. 1914. Über die methoden der korrelationsrechnung in der padagogik und psychologie. *Zeitschrift für Padagogische Psychologie und Experimentelle Padagogik* 15, 114-139, 145-159, 229-242.
- Goodman, L.A. and Kruskal W.H. 1954. Measures of association for cross classifications. *Journal of the American Statistical Association* 49, 733-764.
- Measures of association for cross classifications. II. Further discussion and references. *Journal of the American Statistical Association* 54, 123-163.
- Guttman, L. 1941. An outline of the statistical theory of prediction. In P. Horst, ed. *The Prediction of Personal Adjustment*. New York: Social Science Research Council.
- Kendall, M.G. 1938. A new measure of rank correlation. *Biometrika* 30, 81-93.
- 1948. *Rank Correlation Methods*. London: Charles Griffin.
- Kim, J.O. 1971. Predictive measures of ordinal association. *American Journal of Sociology* 76, 891-907.
- Kruskal, W.H. 1958. Ordinal measures of association. *Journal of the American Statistical Association* 53, 814-861.
- Loether, H.J. and McTavish D.G. 1976. *Descriptive and Inferential Statistics: an Introduction*. Boston: Allyn and Bacon.
- Leik, R.K. and Gove W.R. 1969. The concept and measurement of asymmetric monotonic relationships in sociology. *American Journal of Sociology* 74, 696-709.
- 1971. Integrated approach to measuring association. In H.L. Costner, ed. *Sociological Methodology 1971*. San Francisco: Jossey-Bass.
- Sommers, R.H. 1962. A new asymmetric measure of association for ordinal variables. *American Sociological Review* 27, 799-811.
- Spearman, C. 1904. The proof and measurement of association between two things. *American Journal of Psychology* 15, 72-101.

- Stuart, A. 1953. The estimation and comparison of strengths of association in contingency tables. *Biometrika* **40**, 105-110.
- Wilson, T.P. 1974. Measures of association for bivariate ordinal hypotheses. In H.M. Blalock, ed., *Measurement in the Social Sciences*. Chicago: Aldine-Atherton.