COMMUN. STATIST.-THEOR. METH., A6(14), 1341-1359 (1977)

A COEFFICIENT OF MULTIPLE ASSOCIATION BASED ON RANKS

Alan Agresti

University of Florida
Gainesville, Florida

Key Words & Phrases: Kendall's tau, ordinal variables, multiple correlation; proportional reduction in error.

ABSTRACT

A generalization of Kendall's tau is formulated for describing the association between a dependent variable and a collection of independent variables. The coefficient may be defined in terms of the proportional reduction in prediction errors obtained by predicting the ordering of pairs of observations on the dependent variable based on orderings of the pairs on the independent variables. The coefficient is formulated both for continuous and discrete variables. Approximate large-sample distributions are considered for both cases. Some of the properties of this coefficient are discussed and compared with those of other multiple measures of association based on ranks.

1. INTRODUCTION

We propose a coefficient of multiple rank association $\tau_{Y\cdot \underline{X}} = \tau_{Y\cdot X^{(1)},\ldots,X^{(k)}}$ for describing the association between a dependent variable Y and a set of independent variables

Copyright © 1977 by Marcel Dekker, Inc. All Rights Reserved. Neither this work nor any part may be reproduced or transmitted in any form or by any means, electronic or mechanical, including

$(X^{(1)}, X^{(2)}, \ldots, X^{(k)})$. The coefficient is a generalization of Kendall's tau for two variables, in that it utilizes the orderings for pairs of observations on each of the variables. It may be given a proportional reduction in error interpretation based on predicting pairwise ordering on Y using pairwise orderings on $\{X^{(\nu)}, \nu = 1,2,\ldots,k\}$.

In Section 2, we formulate the coefficient for continuous variables. In Section 3, we consider the case in which there are some tied ranks, or the variables are ordinal categorical in nature. The measure is then defined in terms of all pairs of observations untied with respect to Y and it is seen to have similar properties as in the full-rank (no ties) case. In Section 4, the calculation of the asymptotic sampling distribution of the random sample version $t_{Y\cdot\underline{X}}$ of $\tau_{Y\cdot\underline{X}}$ is discussed. In Section 5, $\tau_{Y\cdot\underline{X}}$ is compared to other multiple rank coefficients which have been proposed.

2. A MULTIPLE TAU COEFFICIENT

The multiple tau coefficient which we define in this section is based on a generalization of the proportional reduction in error interpretation for Kendall's tau (denoted by τ_{YX}). For this bivariate case, let P(C) and P(D) represent the proportions of concordant and discordant pairs of observations, and suppose that there are no tied pairs. If one were to predict at random for each pair of observations whether that pair was concordant or discordant (i.e., if $X_j > X_i$ for the pair (X_i,Y_i) and (X_j,Y_j), then predict $Y_j > Y_i$ with probability ½ and predict $Y_j < Y_i$ with probability ½), the expected proportion of prediction errors would be $(P(C)+P(D))/2 = ½$. If, on the other hand, one knows that $\tau_{YX} > 0$ ($\tau_{YX} < 0$) and predicts concordance (discordance) for every pair, the proportion of errors would be P(D) (P(C)). This results in a proportional reduction in error of $P(C) - P(D) = \tau_{YX}$ $(P(D) - P(C) = |\tau_{YX}|)$.

2.1 Definition of $\tau_{Y\cdot\underline{X}}$

Now, suppose that we wish to describe the association between a dependent variable Y and a collection of independent variables $\underline{X} = (X^{(1)}, \ldots, X^{(k)})$. The variables are required only to be at least ordinal in scale, since just the rankings are utilized. We next construct a coefficient which has predictions of the ordering on Y based on the orderings on the $\{X^{(\nu)}, \nu = 1,\ldots,k\}$ for each pair of observations. We assume that the proportion of pairs tied on any of the variables is zero.

Let $(Y_i, X_i^{(1)}, \ldots, X_i^{(k)})$ and $(Y_j, X_j^{(1)}, \ldots, X_j^{(k)})$ denote the measurements or rankings for a pair (i,j) selected at random. Let

$$S_\nu(i,j) = S[(Y_j - Y_i)(X_j^{(\nu)} - X_i^{(\nu)})], \qquad \nu = 1,\ldots,k, \tag{2.1}$$

where S is the sign function

$$S[u] = \begin{cases} -1, & u < 0 \\ 0, & u = 0 \\ 1, & u > 0. \end{cases}$$

Also, denote $(S_1(i,j),\ldots,S_k(i,j))$ by $\underline{S}(i,j)$, and let

$$P(\underline{\delta}) = P(\delta_1,\ldots,\delta_k) = \Pr\{(i,j) : \underline{S}(i,j) = \underline{\delta}\}. \tag{2.2}$$

For example, $P(1,\ldots,1)$ is the probability that a pair of observations is simultaneously concordant between Y and each $X^{(\nu)}$, $\nu = 1, 2,\ldots,k$. If $\underline{S}(i,j) = \underline{\delta}$ for the pair (i,j), then that pair is called $\underline{Y - X^{(\nu)}\text{ concordant}}$ ($\underline{Y - X^{(\nu)}\text{ discordant}}$) if $\delta_\nu = 1$ ($\delta_\nu = -1$). Let

$$D_k = \{(\delta_1,\ldots,\delta_k) : \delta_\nu = \pm 1, \nu = 1,\ldots,k\}. \tag{2.3}$$

For each element $\underline{\delta}$ of D_k, $P(\underline{\delta}) + P(-\underline{\delta})$ is the probability that a pair has a certain fixed ordering on the $\{X^{(\nu)}\}$, namely

$$X^{(u)} - X^{(w)} \text{ concordant if } \delta_u\delta_w = 1$$
$$X^{(u)} - X^{(w)} \text{ discordant if } \delta_u\delta_w = -1, \qquad 1 \le u \le w \le k.$$

Consider the rule which specifies that for each set of pairs with particular fixed orderings on the $\{X^{(\nu)}\}$, one predicts ordering on Y such that

$$\underline{S}(i,j) = \begin{cases} \underline{\delta} & \text{if } P(\underline{\delta}) > P(-\underline{\delta}) \\ -\underline{\delta} & \text{if } P(\underline{\delta}) < P(-\underline{\delta}) \\ \text{either} & \text{if } P(\underline{\delta}) = P(-\underline{\delta}). \end{cases} \tag{2.4}$$

According to this prediction rule, the probability that a pair is in that set and has ordering on Y incorrectly predicted is $\min(P(\underline{\delta}),P(-\underline{\delta}))$. On the other hand, this probability for a random prediction rule in which $Y - X^{(\nu)}$ concordance or $Y - X^{(\nu)}$ discordance is predicted with probability ½ for each is $(P(\underline{\delta})+P(-\underline{\delta}))/2$. When predictions are considered over all such sets with fixed $\{X^{(\nu)}\}$ orderings, the proportional reduction in error is

$$\tau_{Y\cdot\underline{X}} = \frac{\Sigma_{D_k}(P(\underline{\delta})+P(-\underline{\delta}))/2 - \Sigma_{D_k}\min(P(\underline{\delta}),P(-\underline{\delta}))}{\Sigma_{D_k}(P(\underline{\delta})+P(-\underline{\delta}))/2}$$
$$= \tfrac{1}{2}\Sigma_{D_k}|P(\underline{\delta})-P(-\underline{\delta})|. \tag{2.5}$$

The factor of ½ occurs here and in some subsequent formulas due to the fact that both $|P(\underline{\delta})-P(-\underline{\delta})|$ and $|P(-\underline{\delta})-P(\underline{\delta})|$ occur in these sums when D_k is used as the index set.

Notice that $\tau_{Y\cdot\underline{X}}$ may be expressed as

$$\tau_{Y\cdot\underline{X}} = \tfrac{1}{2}\Sigma_{D_k}(P(\underline{\delta})+P(-\underline{\delta}))|\tau(\underline{\delta})|, \tag{2.6}$$

where

$$\tau(\underline{\delta}) = (P(\underline{\delta})-P(-\underline{\delta}))/(P(\underline{\delta})+P(-\underline{\delta})). \tag{2.7}$$

Hence, $\tau_{Y\cdot\underline{X}}$ is a weighted average of absolute values of Kendall's tau-type measures computed within each set of fixed orderings on the $\{X^{(\nu)}\}$. Since the joint orderings of the $\{X^{(\nu)}\}$ are fixed for

those pairs with $\underline{S} = \underline{\delta}$ or $\underline{S} = -\underline{\delta}$, $|\tau(\underline{\delta})|$ is in fact the absolute value of Kendall's tau between Y and each of the $X^{(\nu)}$ $(\nu = 1,\ldots,k)$, for that set of pairs. Alternatively, let

$$D_M = \{\underline{\delta} : P(\underline{\delta}) > P(-\underline{\delta})\}, \quad D_m = \{\underline{\delta} : P(\underline{\delta}) < P(-\underline{\delta})\}. \tag{2.8}$$

Then, we could rewrite the coefficient as

$$\tau_{Y\cdot\underline{X}} = P_M - P_m = \Sigma_{D_M}(P(\underline{\delta}) - P(-\underline{\delta})), \tag{2.9}$$

where $P_M = \Sigma_{D_M} P(\underline{\delta}) = \Pr(\underline{S}(i,j) \text{ in } D_M)$ and $P_m = \Pr(\underline{S}(i,j) \text{ in } D_m)$ for a randomly selected pair (i,j). We shall refer to the pairs indexed by D_M as those with <u>majority ordering</u> on Y with respect to $\{X^{(\nu)}\}$, and by D_m as those with <u>minority ordering</u> on Y with respect to $\{X^{(\nu)}\}$. Thus $\tau_{Y\cdot\underline{X}}$ is also similar in structure to Kendall's tau in that it may be interpreted as the difference in the probabilities of two types of pairs of observations.

2.2 Properties and Example

We shall next consider some of the properties of $\tau_{Y\cdot\underline{X}}$. It is clear from its definition that $\tau_{Y\cdot\underline{X}}$ is invariant under order-preserving transformations on any of the variables. In the simple bivariate case,

$$\tau_{Y\cdot X} = |P(1) - P(-1)| = |P(C) - P(D)| = |\tau_{YX}|. \tag{2.10}$$

In the trivariate case,

$$\begin{aligned} \tau_{Y\cdot X^{(1)},X^{(2)}} &= |P(1,1) - P(-1,-1)| + |P(1,-1) - P(-1,1)| \\ &= \max\{|(P(1,1)+P(1,-1)) - (P(-1,1)+P(-1,-1))|, \\ &\qquad |(P(1,1)+P(-1,1)) - (P(1,-1)+P(-1,-1))|\} \\ &= \max\{|\tau_{YX^{(1)}}|, |\tau_{YX^{(2)}}|\}. \end{aligned} \tag{2.11}$$

The behavior of $\tau_{Y\cdot\underline{X}}$ becomes less trivial when k exceeds two, as the simultaneous predictive power available from $(X^{(1)},\ldots,X^{(k)})$ may exceed that of the one most strongly associated with Y. In general, it can be shown that $\tau_{Y\cdot X^{(1)},\ldots,X^{(k)}} \le \tau_{Y\cdot X^{(1)},\ldots,X^{(k+1)}}$.

Equality occurs if and only if for each choice of $(\delta_1,\ldots,\delta_k)$, either

$$P(\delta_1,\ldots,\delta_k,1) \geq P(-\delta_1,\ldots,-\delta_k,-1)$$

and

$$P(\delta_1,\ldots,\delta_k,-1) \geq P(-\delta_1,\ldots,-\delta_k,1),$$

or

$$P(\delta_1,\ldots,\delta_k,1) \leq P(-\delta_1,\ldots,-\delta_k,-1)$$

and

$$P(\delta_1,\ldots,\delta_k,-1) \leq P(-\delta_1,\ldots,-\delta_k,1).$$

In particular if $|\tau_{X(\ell)X(k+1)}| = 1$ for some ℓ $(1 \leq \ell \leq k)$, then equality results.

As an example of the computation of $\tau_{Y\cdot\underline{X}}$, we consider some data adapted from Kendall (1970, p. 121). In the Table, Kendall identified the variables Y, $X^{(1)}$, and $X^{(2)}$ with intelligence, mathematical ability, and musical ability, respectively. The Kendall's tau values are $\tau_{YX}(1) = \tau_{YX}(2) = .644$. Now suppose that a variable $X^{(3)}$ is added to the system, with rankings as also given in the Table, for which $\tau_{YX}(3) = -.156$. The 45 pairs of observations may be partitioned into the following sets:

TABLE

Rankings for Numerical Example of $\tau_{Y\cdot\underline{X}}$

Individual	Y	$X^{(1)}$	$X^{(2)}$	$X^{(3)}$
A	1	1	4	1
B	2	4	1	8
C	3	5	3	7
D	4	6	5	6
E	5	2	2	9
F	6	7	6	5
G	7	3	7	10
H	8	9	10	3
I	9	8	9	2
J	10	10	8	4

δ	Pairs with $S(i,j) = \delta$
(1,1,1)	(A,D), (A,F), (A,G), (A,H), (A,I), (A,J), (E,G)
(-1,-1,-1)	(H,I)
(1,1,-1)	(B,C), (B,D), (B,F), (B,H), (B,I), (B,J), (C,D), (C,F), (C,H), (C,I), (C,J), (D,F), (D,H), (D,I), (D,J), (E,F), (E,H), (E,I), (E,J), (F,H), (F,I), (F,J), (G,H), (G,I), (G,J)
(-1,-1,1)	(C,E), (D,E)
(1,-1,1)	(A,B), (A,C), (A,E), (H,J), (I,J)
(-1,1,-1)	None
(1,-1,-1)	None
(-1,1,1)	(B,E), (B,G), (C,G), (D,G), (F,G)

For these pairs, $P_M = P(1,1,1) + P(1,1,-1) + P(1,-1,1) + P(-1,1,1) = 42/45 = .933$ and $P_m = .067$, so that $\tau_{Y\cdot\underline{X}} = .867$. That is, 93.3% of the Y pair orderings may be predicted correctly, which is an 86.7% reduction in error from the 50% correct expected for random predictions.

3. A MULTIPLE TAU COEFFICIENT FOR ORDINAL CATEGORICAL DATA

Tied pairs of observations would typically occur for most systems of variables in the social and behavioral sciences, where variables are commonly measured on ordinal categorical scales. If only a small proportion of pairs of observations are tied on at least one of the variables, one could continue to use $\tau_{Y\cdot\underline{X}}$ as defined in the previous section (tied pairs being ignored in the calculation). However, this results in a reduction in the potential magnitude of the measure which becomes substantial as the proportion of tied pairs increases. For example, if the dependent variable is dichotomous with proportions .2 and .8 of observations in the two categories, the maximum possible value for $\tau_{Y\cdot\underline{X}}$ would be .32 (the proportion of pairs untied on Y) regardless of the distribution of ties among the independent variables.

To permit a maximum value of one and to ensure that the value does not decrease as independent variables are added to the system, one could base the coefficient on those pairs untied on Y. That is, for $\delta_\nu = -1$, 0 or 1, $\nu = 1,\ldots,k$, let

$$P(\underline{\delta}) = \Pr\{(i,j) : S[Y_j - Y_i] \neq 0 \text{ and } \underline{S}(i,j) = \underline{\delta}\} \qquad (3.1)$$

Let P_T denote the probability that a pair is tied with respect to Y. For example, if there are a_o distinct values of Y with proportion p_i at the i-th level, then $P_T = \Sigma_{i=1}^{a_o} p_i^2$. Then, letting

$$D'_k = \{\underline{\delta} : \delta_\nu = -1,\ 0,\ \text{or } +1,\ \nu = 1,\ldots,k\},$$
$$D'_M = \{\underline{\delta} \text{ in } D'_k : P(\underline{\delta}) > P(-\underline{\delta})\}, \qquad (3.2)$$

we define

$$\tau_{Y\cdot\underline{X}} = \tfrac{1}{2}\Sigma_{D'_k} |P(\underline{\delta}) - P(-\underline{\delta})|/(1 - P_T)$$
$$= \Sigma_{D'_M} (P(\underline{\delta}) - P(-\underline{\delta}))/(1 - P_T). \qquad (3.3)$$

We observe that $\tau_{Y\cdot\underline{X}}$ may be expressed as

$$\tau_{Y\cdot\underline{X}} = \{\tfrac{1}{2}(1 - P_T) - \tfrac{1}{2}\Sigma_{D'_k} \min(P(\underline{\delta}),P(-\underline{\delta}))\}/\{\tfrac{1}{2}(1 - P_T)\}. \qquad (3.4)$$

That is, $\tau_{Y\cdot\underline{X}}$ is the proportional reduction in error of predictions of the ordering on Y (for those pairs untied on Y) obtained by predicting majority ordering relative to predicting order randomly. Alternatively, $\tau_{Y\cdot\underline{X}}$ may be expressed as

$$\tau_{Y\cdot\underline{X}} = \Sigma_{D'_k} \lambda(\underline{\delta}) \frac{|P(\underline{\delta}) - P(-\underline{\delta})|}{P(\underline{\delta}) + P(-\underline{\delta})}$$
$$= \tfrac{1}{2}\Sigma_{D'_k} (\lambda(\underline{\delta}) + \lambda(-\underline{\delta}))|\tau(\underline{\delta})|, \qquad (3.5)$$

where $\lambda(\underline{\delta}) = P(\underline{\delta})/(1 - P_T)$ is the proportion of the pairs of observations untied on Y for which $\underline{S}(i,j) = \underline{\delta}$. Hence, $\tau_{Y\cdot X}$ may be interpreted as a weighted average of the absolute values of Kendall's tau-type measures within each fixed set of orderings on the $\{X^{(\nu)}\}$. Here, $|\tau(\underline{\delta})|$ is the absolute value of Kendall's tau between Y and each of the $X^{(\nu)}$ such that $\delta_\nu \neq 0$, within the set of pairs for which $\underline{S} = \underline{\delta}$ or $\underline{S} = -\underline{\delta}$.

Clearly, $\tau_{Y\cdot\underline{X}}$ is invariant under strictly order preserving transformations on any of the variables. When there are no tied pairs with respect to any of the variables, $\tau_{Y\cdot\underline{X}}$ reduces to the coefficient discussed in Section 2, so we have used the same symbol. In the bivariate case, $\tau_{Y\cdot X}$ reduces to the absolute value of Somers' d_{XY} (see Somers (1962)), a well-known asymmetric ordinal measure of association. When $k = 2$, it is likely to be not much larger than $\max(\tau_{Y\cdot X(1)},\tau_{Y\cdot X(2)})$, but there is not necessarily equality here due to the additional contribution in the numerator of pairs tied on $X^{(1)}$ but not on $X^{(2)}$ and Y, or of pairs tied on $X^{(2)}$ but not on $X^{(1)}$ and Y. Again, though, $\tau_{Y\cdot\underline{X}}$ is of primary interest when $k \geq 3$.

It can be shown that $\tau_{Y\cdot X(1),\ldots,X(k)} \leq \tau_{Y\cdot X(1),\ldots,X(k+1)}$, since the denominator remains constant and the numerator can not decrease when a variable is added to the system. Equality results if and only if the introduction of $X^{(k+1)}$ does not alter the predictions of $Y - X^{(\nu)}$ concordance or discordance ($\nu = 1,\ldots k$), and out of those pairs tied on all $X^{(\nu)}$, ($\nu = 1,\ldots,k$), there is the same proportion of $Y - X^{(k+1)}$ concordant and $Y - X^{(k+1)}$ discordant pairs.

4. SAMPLING DISTRIBUTIONS

When the value of any measure is computed from some sample, one would usually be interested in making inferences about the value of the measure for some associated population. We now consider large sample approximations for the first two moments and the distributions of the coefficients under random sampling, for the case when all pairs untied on Y and at least one $X^{(\nu)}$ are of majority or minority ordering. From these distributions, large sample approximate confidence intervals may be formulated for $\tau_{Y\cdot\underline{X}}$.

4.1 Inference - Continuous Variables

Suppose at first that there is probability zero of a tied pair with respect to any of the variables. For all $\underline{\delta}$ in D_k, let $\hat{P}(\underline{\delta})$ be the sample proportion of pairs for which $\underline{S}(i,j) = \underline{\delta}$, and let

$$\hat{D}_M = \{\underline{\delta} : \hat{P}(\underline{\delta}) > \hat{P}(-\underline{\delta})\}$$
$$\hat{D}_m = \{\underline{\delta} : \hat{P}(\underline{\delta}) < \hat{P}(-\underline{\delta})\}. \qquad (4.1)$$

Then the sample value $t_{Y\cdot\underline{X}}$ of $\tau_{Y\cdot\underline{X}}$ may be written as

$$t_{Y\cdot\underline{X}} = 2\Sigma_{i=1}^{n} \Sigma_{j=i+1}^{n} v_{ij}/n(n-1), \qquad (4.2)$$

where

$$v_{ij} = 1 \text{ if } \underline{S}(i,j) \text{ in } \hat{D}_M$$
$$= -1 \text{ if } \underline{S}(i,j) \text{ in } \hat{D}_m$$
$$= 0 \text{ otherwise.}$$

Now,

$$Et_{Y\cdot\underline{X}} = Ev_{ij} = \Pr(\underline{S}(i,j) \text{ in } \hat{D}_M) - \Pr(\underline{S}(i,j) \text{ in } \hat{D}_m)$$
$$= \Sigma_{D_k} P(\underline{\delta})[\Pr(\hat{D}_M \text{ contains } \underline{\delta}) - \Pr(\hat{D}_m \text{ contains } \underline{\delta})].$$

Notice that $t_{Y\cdot\underline{X}}$ is not in general unbiased. If we assume, however, that

$$|P(\underline{\delta}) - P(-\underline{\delta})| > 0 \text{ for all } \underline{\delta} \text{ in } D_k, \qquad (4.3)$$

then $\Pr(\hat{P}(\underline{\delta}) > \hat{P}(-\underline{\delta}) \,|\, P(\underline{\delta}) > P(-\underline{\delta})) \to 1$ as $n \to \infty$, so that

$$\Pr(\hat{D}_M \equiv D_M \text{ and } \hat{D}_m \equiv D_m) \to 1 \text{ as } n \to \infty. \qquad (4.4)$$

It follows then that as $n \to \infty$,

$$Et_{Y\cdot\underline{X}} \to \Sigma_{D_M} P(\underline{\delta}) - \Sigma_{D_m} P(\underline{\delta}) = \tau_{Y\cdot\underline{X}}. \qquad (4.5)$$

For asymptotic purposes, we shall act as if $\hat{D}_M \equiv D_M$ and $\hat{D}_m \equiv D_m$, appealing to the same argument used in a lemma by Goodman and Kruskal (1963, p. 357). In other words, for sufficiently large n, $t_{Y\cdot\underline{X}}$ equals

$$\Sigma_{D_M} (\hat{P}(\underline{\delta}) - \hat{P}(-\underline{\delta})) = \hat{P}_M - \hat{P}_m. \qquad (4.6)$$

The statistical behavior of this coefficient is similar to that of the absolute value of Kendall's tau, in terms of being an (asymptotically) unbiased estimate of the difference between two proportions - the proportions of pairs of majority ordering and pairs of minority ordering on Y with respect to $\{X^{(\nu)}\}$. Hence, it seems reasonable to apply a derivation for the distribution of Kendall's tau (e.g., see Noether (1967, Ch. 10)), with slight modifications, to get a large sample distribution for $t_{Y\cdot\underline{X}}$.

For three observations chosen at random, let

$$P_{MM} = \Pr(\underline{S}(1,2) \text{ and } \underline{S}(1,3) \text{ both in } D_M)$$
$$P_{mm} = \Pr(\underline{S}(1,2) \text{ and } \underline{S}(1,3) \text{ both in } D_m)$$
$$P_{Mm} = \Pr(\underline{S}(1,2) \text{ in } D_M,\ \underline{S}(1,3) \text{ in } D_m)$$
$$P_{mM} = \Pr(\underline{S}(1,2) \text{ in } D_m,\ \underline{S}(1,3) \text{ in } D_M). \qquad (4.7)$$

Using the same argument as for $E v_{ij}$, it can be seen that

$$E v_{ij}^2 \to P_M + P_m = 1 \text{ as } n \to \infty,$$

and

$$E v_{ij} v_{i\ell} \to P_{MM} + P_{mm} - P_{Mm} - P_{mM} \text{ as } n \to \infty.$$

Now,

$$\text{Var}(t_{Y\cdot\underline{X}}) = \left(\frac{2}{n(n-1)}\right)^2 \left[\sum_{i<j} \text{var}(v_{ij}) + \sum_{i<j} \sum_{\substack{i'<j' \\ (i\neq i' \text{ or } j\neq j')}} \text{cov}(v_{ij}, v_{i'j'})\right]$$
$$= \left(\frac{2}{n(n-1)}\right)^2 \left[\binom{n}{2}\text{var}(v_{12}) + 6\binom{n}{3}\text{cov}(v_{12}, v_{13})\right] \qquad (4.8)$$

by symmetry. Using the fact that $P_{Mm} + P_{MM} = P_M$, $P_{mM} + P_{mm} = P_m$, $P_{mM} = P_{Mm}$, we see that for large n the variance of $\sqrt{n}\, t_{Y\cdot\underline{X}}$ is approximately

$$\sigma^2 = 16(P_{MM} - P_M)^2. \tag{4.9}$$

In addition, it seems reasonable based on representation (4.2) to conjecture that the asymptotic distribution of $t_{Y\cdot\underline{X}}$ is normal (at least for a broad class of underlying distributions) so that

$$\sqrt{n}(t_{Y\cdot\underline{X}} - \tau_{Y\cdot\underline{X}}) \xrightarrow{d} N(0,\sigma^2), \tag{4.10}$$

under assumption (4.3). In practice, one would probably not know σ. However, when (4.10) holds,

$$\sqrt{n}(t_{Y\cdot\underline{X}} - \tau_{Y\cdot\underline{X}})/\hat{\sigma} \xrightarrow{d} N(0,1),$$

where $\hat{\sigma}$ is a consistent estimate of σ, by Slutzky's Theorem (see Goodman and Kruskal (1963,p. 356)). For example, one could substitute in the formula for σ^2 the sample values

$$\hat{P}_M = \Sigma_{i=1}^n M_i/n(n-1) = \Sigma_{\hat{D}_M} \hat{P}(\underline{\delta})$$
$$\hat{P}_{MM} = \Sigma_{i=1}^n M_i(M_i - 1)/n(n-1)(n-2), \tag{4.11}$$

where M_i is the number of pairs (i,j) containing the i-th observation for which $\underline{S}(i,j)$ is in D_M. Then, an approximate large-sample $100(1-\alpha)\%$ confidence interval for $\tau_{Y\cdot\underline{X}}$ is given by $t_{Y\cdot\underline{X}} \pm Z_{\alpha/2}\hat{\sigma}/\sqrt{n}$.

If there are some tied pairs, but not enough to treat the data using categorical techniques, then a correction factor can be introduced into the asymptotic variance of $t_{Y\cdot\underline{X}}$, yielding a formula analogous to Noether's (10.8). That is

$$\mathrm{Var}\sqrt{n}(t_{Y\cdot\underline{X}} - \tau_{Y\cdot\underline{X}}) \xrightarrow[n\to\infty]{} 4(P_{MM} + P_{mm} - P_{Mm} - P_{mM}) - 4(P_M - P_m)^2, \tag{4.12}$$

which reduces to the formula for σ^2 in (4.9) when there are no tied pairs. The sample values of P_{mm}, P_{mM}, and P_{Mm} are

$$\hat{P}_{mm} = \Sigma_{i=1}^n m_i(m_i - 1)/n(n-1)(n-2),$$
$$\hat{P}_{mM} = \hat{P}_{Mm} = \Sigma_{i=1}^n m_i M_i/n(n-1)(n-2), \tag{4.13}$$

where m_i is the number of pairs containing the i-th observation for which $\underline{S}(i,j)$ is in $\hat{D}_m$.

4.2 Inference - Ordinal Categorical Variables

Suppose now that there are a_0 categories for Y, a_1 categories for $X^{(1)}$, ..., a_k categories for $X^{(k)}$ and let $p_{i_0\cdots i_k}$ denote the probability (under random sampling) that an observation is in category i_0 of Y, i_1 of $X^{(1)}$, ..., i_k of $X^{(k)}$. We shall only discuss here the situation in which there is full random sampling over the entire multinomial classification (i.e., none of the marginal distributions are treated as fixed). We assume that

$$|P(\underline{\delta}) - P(-\underline{\delta})| > 0 \text{ for all } \underline{\delta} \text{ in } D_k' - \{\underline{0}\}. \tag{4.14}$$

Let $\{\hat{p}_{i_0\cdots i_k}\}$ denote the sample proportions (the m.ℓ.e.'s) corresponding to $\{p_{i_0,\ldots,i_k}\}$. Also, let $\{\hat{P}(\underline{\delta})\}$ and $\{\hat{p}_i\}$ be the same functions of the $\{\hat{p}_{i_0\cdots i_k}\}$ that $\{P(\underline{\delta})\}$ and $\{p_i\}$ are of the $\{p_{i_0\cdots i_k}\}$. For example, for a 2 × 2 × 2 × 2 cross-classification, $\hat{P}(1,1,1) = 2\hat{p}_{1111}\hat{p}_{2222}$. Then, the random sample version $t_{Y\cdot\underline{X}}$ can be written as

$$t_{Y\cdot\underline{X}} = \Sigma_{\hat{D}_M'}(\hat{P}(\underline{\delta}) - \hat{P}(-\underline{\delta}))/(1 - \Sigma_{i=1}^{a_0} \hat{p}_i^2), \tag{4.15}$$

where $\hat{D}_M'$ is the sample version of D_M'. As $n \to \infty$,

$$\Pr(\hat{D}'_M \equiv D'_M) \to 1, \quad (4.16)$$

since the $\{\hat{p}_{i_0\cdots i_k}\}$ are consistent estimates of the $\{p_{i_0\cdots i_k}\}$. For asymptotic arguments, then, we shall again treat D'_M as known. Thus, for large n,

$$\sqrt{n}(t_{Y\cdot \underline{X}} - \tau_{Y\cdot \underline{X}}) = \sqrt{n}\left[\frac{\Sigma_{D'_M}(\hat{P}(\underline{\delta}) - \hat{P}(-\underline{\delta}))}{1 - \Sigma_{i=1}^{a_0} \hat{p}_i^2} - \frac{\Sigma_{D'_M}(P(\underline{\delta}) - P(-\underline{\delta}))}{1 - \Sigma_{i=1}^{a_0} p_i^2}\right], \quad (4.17)$$

which has asymptotically the same distribution as

$$\sqrt{n}\left[\frac{(1-\Sigma_{i=1}^{a_0} p_i^2)\Sigma_{D_M}(\hat{P}(\underline{\delta})-\hat{P}(-\underline{\delta})) - (1-\Sigma_{i=1}^{a_0}\hat{p}_i^2)\Sigma_{D_M}(P(\underline{\delta})-P(-\underline{\delta}))}{(1-\Sigma_{i=1}^{a_0} p_i^2)^2}\right]. \quad (4.18)$$

This last expression is a continuously differentiable function of $\{\hat{p}_{i_0\cdots i_k}\}$ in a neighborhood of $\{p_{i_0\cdots i_k}\}$. Using the fact that the $\{\sqrt{n}(\hat{p}_{i_0\cdots i_k} - p_{i_0\cdots i_k})\}$ are jointly asymptotically normally distributed about 0 with

$$\mathrm{Var}\left[\sqrt{n}(\hat{p}_{i_0\cdots i_k} - p_{i_0\cdots i_k})\right] = p_{i_0\cdots i_k}(1 - p_{i_0\cdots i_k}),$$

and

$$\mathrm{Cov}\left[\sqrt{n}(\hat{p}_{i_0\cdots i_k} - p_{i_0\cdots i_k}), \sqrt{n}(\hat{p}_{j_0\cdots j_k} - p_{j_0\cdots j_k})\right]$$

$$= -p_{i_0\cdots i_k} p_{j_0\cdots j_k},$$

it then follows that $\sqrt{n}(t_{Y\cdot \underline{X}} - \tau_{Y\cdot \underline{X}})$ is asymptotically normally distributed about zero with variance

$$\sigma^2 = \frac{(\Sigma_{i_0}\cdots\Sigma_{i_k} p_{i_0\cdots i_k}\phi^2_{i_0\cdots i_k}) - (\Sigma_{i_0}\cdots\Sigma_{i_k} p_{i_0\cdots i_k}\phi_{i_0\cdots i_k})^2}{(1-\Sigma_{i=1}^{a_0} p_i^2)^4}, \quad (4.19)$$

where

$$\phi_{i_0\cdots i_k} = (1-\Sigma_{i=1}^{a_0} p_i^2)\Sigma_{D'_M}\left(\frac{\partial \hat{P}(\underline{\delta})}{\partial \hat{p}_{i_0\cdots i_k}} - \frac{\partial \hat{P}(-\underline{\delta})}{\partial \hat{p}_{i_0\cdots i_k}}\right)\Bigg|_{\{p_{i_0\cdots i_k}\}}$$

$$+ 2p_{i_0}\Sigma_{D'_M}(P(\underline{\delta}) - P(-\underline{\delta})). \quad (4.20)$$

Notice that the asymptotic variance depends on which of $P(\underline{\delta})$ and $P(-\underline{\delta})$ is the maximum, for all $\underline{\delta}$ in D'_k. In practice, σ^2 may be replaced by its maximum likelihood estimate $\hat{\sigma}^2$ (the same formula with the substitution of the sample proportions $\{\hat{p}_{i_0\cdots i_k}\}$). Then, $t_{Y\cdot \underline{X}} \pm Z_{\alpha/2}\hat{\sigma}/\sqrt{n}$ is a large-sample approximate $100(1-\alpha)\%$ confidence interval for $\tau_{Y\cdot \underline{X}}$. The calculations are very cumbersome for a large number of independent variables or a fine classification of variables, but can be handled very simply for many cross-classifications using a computer program.

5. A COMPARISON WITH OTHER MULTIPLE ORDINAL COEFFICIENTS

The best known and most commonly used measure of multiple rank association, apparently first proposed by Moran (1951), is

$$\tau_M^2 = 1-(1-\tau^2_{YX(1)})(1-\tau^2_{YX(2)\cdot X(1)})(1-\tau^2_{YX(3)\cdot X(1)X(2)})\cdots$$

$$\cdots(1-\tau^2_{YX(k)\cdot X(1),\ldots,X(k-1)}), \quad (5.1)$$

where $\tau_{YX(1)}$ is Kendall's tau between Y and $X^{(1)}$,

$$\tau_{YX(2)\cdot X(1)} = (\tau_{YX(2)} - \tau_{YX(1)}\tau_{X(1)X(2)})/[(1-\tau^2_{YX(1)})(1-\tau^2_{X(1)X(2)})]^{\frac{1}{2}}$$

is Kendall's partial tau of order 1, etc. The formula for τ_M^2 is analogous to the one for the coefficient of determination with Kendall's tau substituting for the Pearson correlation. In fact, this coefficient is the same as the coefficient of determination corresponding to a linear model using sign scores based on the n(n - 1) ordered pairs of observations. That is, in using least squares to fit the equation

$$S[Y_j - Y_i] = b_1 S[X_j^{(1)} - X_i^{(1)}] + b_2 S[X_j^{(2)} - X_i^{(2)}] + \cdots + b_k S[X_j^{(k)} - X_i^{(k)}] \quad (5.2)$$

to the observed pair scores, the partial taus and multiple τ_M are the results of applying the formulae for the Pearson partial and multiple correlation.

Based on this construction, τ_M^2 can be given the usual proportional reduction in error interpretation (based on the sum of squared prediction errors) in terms of predicting Y pair scores using the linear function of $\{X^{(\nu)}\}$ pair scores, as compared to predicting them randomly. The reader should see the articles by Hawkes (1971) and Ploch (1974) for more complete expositions concerning the multivariate analysis of ranked data by means of a system including these partial and multiple measures.

A related approach is to formulate a multiple correlation measure corresponding to the linear model in which the rank on Y is modelled as a linear function of the ranks on the $\{X^{(\nu)}\}$. When $k = 1$, this gives the Spearman's rho rank correlation coefficient. For general k, this measure is the same function of the pairwise Spearman's rhos that the multiple correlation coefficient is of pairwise Pearson correlations. This coefficient is implied as a special case of a measure of association in the general approach to rank tests of independence given by Puri and Sen (1971, Ch. 8).

When there are tied pairs on at least one of the variables, the zero-order taus which constitute τ_M^2 become τ_b's (Kendall (1970, p. 35)), and (5.1) is used with the substitution of that generalization of tau for tied ranks (see Hawkes (1971), Ploch, (1974)). Alternatively, for ordinal categorical variables, Morris (1970) proposed using the Somers d_{XY} bivariate measure on the single table composed of the Y classification crossed with a classification based on all possible combinations of the categories of the independent variables (one category from each variable).

The obvious difficulty with this latter approach is in deciding how to order these newly created categories, especially if there are several independent variables or several levels for each one. There is no unique way to define the ordering of this new classification, and of course different choices for this ordering could lead to drastically different values of the measure.

In many situations, the coefficient τ_M would be adequate by itself for describing the extent of multiple association. One should be aware, however, that its behavior need not strictly parallel that of the multiple correlation $R_{Y \cdot X^{(1)},\ldots,X^{(k)}}$, for a given set of variables. For example, suppose that we are considering the population values when the joint distribution of $(Y, X^{(1)}, X^{(2)})$ is a trivariate normal. Then, using the fact that Kendall's tau is related to the Pearson correlation ρ for a bivariate normal distribution by

$$\tau = \frac{2}{\pi} \sin^{-1} \rho ,$$

it can be seen using differential calculus that if $\rho_{YX(2) \cdot X(1)} = 0$, then $\tau_{YX(2) \cdot X(1)} = 0$ only in the trivial case when $\rho_{X(1)X(2)}$ or $\rho_{YX(2)}$ equals -1, 0 or 1. Thus, in general in the normal case with a spuriously related independent variable $X^{(2)}$, $\tau_{YX(2) \cdot X(1)} \neq 0$ and hence $\tau_M > |\tau_{YX(1)}|$ even though $R_{Y \cdot X(1)X(2)} = |\rho_{YX(1)}|$. Also, one could have a trivariate normal system with $\rho_{YX(2) \cdot X(1)} \neq 0$ but $\tau_{YX(2) \cdot X(1)} = 0$, so that $\tau_M = |\tau_{YX(1)}|$, though $R_{Y \cdot X(1), X(2)} > |\rho_{YX(1)}|$. Similar deviations, though not as extreme, tend to occur for the Spearman-type multiple correlation based on ranks.

The multiple rank coefficient τ_M has other deficiencies, as well. For example, its sampling distribution is too complex to allow the formation of confidence intervals for the population value, even for large samples. The proportional reduction in error interpretation for τ_M^2 is somewhat artificial, in the sense that the predicted Y pair scores obtained from (5.2) are not -1, 0 or +1 in general, and in fact need not even be between -1 and +1.

In addition, when $k \geq 3$, τ_M is only adequate if the simple additive relationship given by (5.2) is the appropriate model for the relationship among the pair scores.

It is the intent of this article to introduce the coefficient $\tau_{Y \cdot \underline{X}}$ as a supplementary measure for describing the rank-order association between Y and $(X^{(1)},\ldots,X^{(k)})$. It should be clear that τ_M and $\tau_{Y \cdot \underline{X}}$ measure two somewhat different facets of association, and thus should not be viewed as competitors. Among the nice features of $\tau_{Y \cdot \underline{X}}$ are a simple interpretation and the fact that it is not just appropriate for a certain type of relationship among the pair scores. However, we have also seen ways in which the behavior of $\tau_{Y \cdot \underline{X}}$ is unlike that of the multiple correlation, probably even more so than τ_M. For example, $\tau_{Y \cdot X(1),\ldots,X(k)}$ may equal $\tau_{Y \cdot X(1),\ldots,X(k+1)}$ when there is a nonzero partial association between Y and $X^{(k+1)}$ given $X^{(1)},\ldots,X^{(k)}$, in the sense that the value of tau for the joint distribution of Y and $X^{(k+1)}$ conditional on $X^{(1)},\ldots,X^{(k)}$ is nonzero. This occurs when, given the ordering of a pair of observations on $X^{(1)},\ldots,X^{(k)}$, no further predictive power (as defined in Section 2) is obtained from knowledge of the ordering on $X^{(k+1)}$.

Of course, similar situations occur in many other contexts in measuring association. For example, the measure lambda which is commonly used for describing association between nominal variables (Goodman and Kruskal (1954)) is based on prediction of the category of the dependent variable with modal frequency. It will be zero even though two variables are statistically dependent, if there is no reduction in prediction errors due to knowledge of the categorization on the independent variable. In conclusion, it should be emphasized that these deficiencies do not invalidate the use of these coefficients. Each is useful in the appropriate circumstances as long as one understands the prediction rule and definition of error for the coefficient, and interprets its value according to these or other available descriptions. In this vein, we propose that a coefficient such as $\tau_{Y \cdot \underline{X}}$ should be considered as a complementary means of describing multiple rank association.

ACKNOWLEDGEMENTS

I would like to thank the referee and Mr. Christian Loschcke for several very helpful comments.

BIBLIOGRAPHY

Goodman, L. A. & Kruskal, W. H. (1954). Measures of association for cross classifications. J. Amer. Statist. Assoc. 49, 732-64.

Goodman, L. A. & Kruskal, W. H. (1963). Measures of association for cross classifications III: Approximate sampling theory. J. Amer. Statist. Assoc. 58, 310-64.

Hawkes, R. K. (1971). The multivariate analysis of ordinal measures. Amer. J. Sociol. 76, 908-26.

Kendall, M. G. (1970). Rank Correlation Methods. (4th ed.). London: Charles W. Griffin and Co., Ltd.

Moran, P. A. P. (1951). Partial and multiple rank correlation. Biometrika 38, 26-32.

Morris, R. N. (1970). Multiple correlation and ordinally scaled data. Social Forces 48, 299-311.

Noether, G. E. (1967). Elements of Nonparametric Statistics. New York: John Wiley & Sons, Inc.

Ploch, D. R. (1974). Ordinal measures of association and the general linear model. Measurement in the Social Sciences (Hubert M. Blalock, Ed.). Chicago: Aldine-Atherton, Ch. 12.

Puri, M. L. and Sen, K. S. (1971). Nonparametric Methods in Multivariate Analysis. New York: John Wiley & Sons, Inc.

Somers, R. H. (1962). A new asymmetric measure of association for ordinal variables. Amer. Sociol. Rev. 27, 799-811.

Received October, 1976; revised February and July, 1977.

Recommended by R. L. Scheaffer, University of Florida, Gainesville.

Refereed by Dana Quade, The University of North Carolina at Chapel Hill.