PSYCHOMETRIKA—VOL. 41, NO. 2
JUNE, 1976

A FURTHER NOTE ON FREEMAN'S MEASURE OF ASSOCIATION

LINTON C. FREEMAN

LEHIGH UNIVERSITY

This note extends and elaborates Hubert's attempt to provide an interpretation of Freeman's measure of association, θ. The θ measure is used in a contingency table when observations are ordered on one variable and unordered on the other. No attempt is made explore the distribution of θ.

In a recent note on θ, Freeman's measure of association, Hubert [1974] suggested that the interpretation of the statistic given in Freeman's [1965] introductory text left something to be desired. He proceeded, therefore, to demonstrate that θ could be understood in terms of its relationship to Sommers' [1962] asymmetric measure of association.

Hubert's criticism was, in my view, well founded. Moreover, his demonstration of the relationship between θ and Sommers' statistic represents an important step in specifying the meaning of θ. This note is intended to extend Hubert's treatment by suggesting an alternative formulation of θ.

Suppose that we have k univariate populations or distributions on an ordered variable X. The distributions may be either discrete or continuous. We are interested here in the degree to which the order on X of pairs of observations from different populations can be predicted from knowledge about population membership. Thus the measure, θ, is in some sense a non-metric analogue of the correlation ratio, η. Our ability to predict correctly depends upon the degree to which observations in a given population are consistently higher (or lower) than those in other populations. If the observations in a population are equal to those in another, or if half of them are higher and half lower, we will be unable to predict order from population membership.

Consider two hypothetical independent observations, one from each of two of the populations X_i, X_j; $i \neq j$. We wish to predict the ordering of X_i, X_j (either $X_i < X_j$ or $X_j < X_i$). If we do not know i and j and their distributions we can do no better than to flip a fair coin. This will lead us to choose $X_i < X_j$, with a probability of 1/2 and $X_j < X_i$ with the same probability.

Suppose that the actual probabilities are

$$\Pr\,(X_i < X_j) = L_{ij}\ ,$$

$$\Pr\,(X_j < X_i) = L_{ji}\ ,$$

and

$$\Pr\,(X_j = X_i) = T_{ij}\ .$$

Requests for reprints should be sent to L. C. Freeman, Department of Social Relations, Price Hall, Bldg. 40, Lehigh University, Bethlehem, Pa. 18015.
Key words: theta, ordinal association.

Then

$$L_{ij} + L_{ji} + T_{ij} = 1.$$

The probability of an erroneous prediction will be $L_{ji} + T_{ij}$ if we pick $X_i < X_j$, and $L_{ij} + T_{ij}$ if we pick $X_j < X_i$. Thus, the probability of incorrect prediction of order, if we predict by flipping a fair coin, is

$$1/2(L_{ij} + T_{ij}) + 1/2(L_{ji} + T_{ij}) = 1/2(1 + T_{ij}).$$

Since we have k populations, there will be $k(k - 1)/2$ unordered pairs which may be chosen for comparisons such as the above. This is simply the number of combinations of elements which can be chosen two at a time from a population of k elements, $\binom{k}{2}$. If we now average the probability of error uniformly over these pairs (corresponding to a choice of two populations from k, at random without replacement) the average probability of error, predicting from ignorance, is

$$E_I = \frac{\sum\sum_{i<j} (1 + T_{ij})}{k(k - 1)}.$$

For a specific pair of distributions, $i \neq j$, suppose now that we know the two distributions, in particular the values of L_{ij}, L_{ji}, and T_{ij}. Let us agree to predict $X_i < X_j$ or $X_j < X_i$ according to which has the greater probability. In this case, the probability of an erroneous prediction of order is equal to the smaller of the two values $L_{ij} + T_{ij}$ or $L_{ji} + T_{ij}$. The smaller value is

$$1/2\left[1 - \left|\left(L_{ij} + \frac{T_{ij}}{2}\right) - \left(L_{ji} + \frac{T_{ij}}{2}\right)\right|\right] + \frac{T_{ij}}{2}.$$

If we let $\Delta_{ij} = |L_{ij} - L_{ji}|$, the expression simplifies to $1/2\ (1 - \Delta_{ij} + T_{ij})$, the probability of erroneous prediction of order when we predict optimally from knowledge of the distributions of X_i and X_j.

The average probability of error uniformly over the $1/2\ k(k - 1)$ choices of i, j $(i < j)$ based upon optimal prediction of order from knowledge of the distributions is

$$\begin{aligned} E_K &= \frac{1}{1/2k(k-1)}\left[\sum\sum_{i<j} 1/2(1 - \Delta_{ij} + T_{ij})\right] \\ &= \frac{1}{k(k-1)}\left[\sum\sum_{i<j} (1 - \Delta_{ij} + T_{ij})\right] \\ &= \frac{\sum\sum_{i<j} (1 - \Delta_{ij} + T_{ij})}{k(k-1)}. \end{aligned}$$

In line with Costner's [1965] suggestion, θ can now be defined as the decrease in the probability of erroneous prediction of order, with optimal prediction of order based on knowledge of the i and j distributions, relative to the probability of error under picking by tossing a fair coin.

$$
\begin{aligned}
\theta &= \frac{E_I - E_K}{E_I} \\
&= \frac{\dfrac{\sum\sum_{i<j} (1 + T'_{ij})}{k(k-1)} - \dfrac{\sum\sum_{i<j} (1 - \Delta_{ij} + T_{ij})}{k(k-1)}}{\dfrac{\sum\sum_{i<j} (1 + T_{ij})}{k(k-1)}} \\
&= \frac{\sum\sum_{i<j} \Delta_{ij}}{\sum\sum_{i<j} (1 + T_{ij})}.
\end{aligned}
$$

Thus θ is a measure of association between a k-class nominal scale and an ordered variable. It is based upon systematic regularities in order between pairs of populations in different classes. The measure θ has the following properties:

(1) $\theta = 0$ if and only if $L_{ij} = L_{ji}$ for all $i \neq j$. (Note that this includes the case where $T_{ij} = 1$.) In such cases the k populations exhibit stochastic equality and we commit as much error in predicting order optimally on the basis of population distribution as we do under chance prediction. It should be observed that only stochastic equality and not identity of the distributions is required in this case.

(2) $\theta = 1$ if and only if $T_{ij} = 0$ and $L_{ij} = 1$ or 0 for all pairs $i \neq j$. When $\theta = 1$ there is no overlap and there are no ties in the rankings of the k populations. The observations in each class are *consistently* higher or lower than those in each of the other classes. Thus, the order of any pair of observations from different populations can be predicted without error from a knowledge of the population distributions.

When the distribution of X is continuous the expression for θ can be simplified. Since $T_{ij} = 0$ for continuous distributions,

$$\theta = \frac{\sum\sum_{i<j} \Delta_{ij}}{1/2[k(k-1)]}.$$

This suggests an alternative interpretation of θ. For continuous distributions, θ is the average absolute difference between the probability that an observation, X_i, from one population is less than an observation, X_j, from another population, and the probability that X_j is less than X_i. The earlier interpretation, of course, is also true for this continuous case.

REFERENCES

Costner, H. L. Criteria for measures of association. *American Sociological Review*, 1965, **30**, 341–353.

Freeman, L. C. *Elementary applied statistics.* New York: Wiley, 1965.

Hubert, L. A note on Freeman's measure of association for relating on ordered to an unordered factor. *Psychometrika*, 1974, **39**, 517–520.

Sommers, R. H. A new asymmetric measure of association for ordinal variables. *American Sociological Review*, 1962, **27**, 799–811.

Manuscript received 3/26/75
Final version received 9/9/75