A NOTE ON FREEMAN'S MEASURE OF ASSOCIATION FOR RELATING AN ORDERED TO AN UNORDERED FACTOR

LAWRENCE HUBERT
UNIVERSITY OF WISCONSIN

Within the context of a contingency table, this note describes the relationship between Freeman's measure of association $\theta$ and the asymmetric association measures developed by Somers. The $\theta$ coefficient is appropriate for a contingency table in which the levels of one factor are ordered and the levels of the other factor are unordered; the indices defined by Somers are usually used when the levels of both factors are ordered and one is assumed to be the independent factor.

One of the major concerns of quantitative methodologists in recent years has been the development of indices of association in contingency tables that have operational significance. The papers by Goodman and Kruskal [1954, 1959, 1963, 1972] are probably the best indication of this concern, but other substantively oriented statisticians have been active in this area of research as well, for instance, Somers [1962], Elashoff [1971] and Särndal [1974]. For contingency tables in which the rows and columns are both inherently ordered or both unordered, the Goodman-Kruskal framework is rather complete. Apparently, however, no serious attempt has been made to develop a correspondingly elegant extension to the mixed contingency table in which, say, the rows are unordered but the columns are ordered, i.e., to the nominal-ordinal case. Särndal [1974] presents some recent work in this direction that should be consulted by the interested reader.

The purpose of this note is to reformulate one suggestion given by Freeman [1965] for a nominal-ordinal measure of association within the clear framework Somers uses to present an asymmetric version of the Goodman-Kruskal $\gamma$ coefficient. Since no concept of a sampling distribution will be discussed below, we assume that our contingency table includes the total population of interest.

Suppose that $n$ objects can be cross classified by two factors $A$ and $B$; for notational purposes $A$ will be the row factor and $B$ the column factor, with $n_{ij}$ objects in row $i$ and column $j$, $1 \leq i \leq I$, $1 \leq j \leq J$. Column and row marginal frequencies will be denoted by $n_{.i}$ and $n_{.j}$, respectively, for the same range of subscripts. Given this basic data, Somers defines the following classification of all possible $\binom{n}{2}$ object pairs, assuming that the rows and
columns are inherently ordered consistently with respect to the subscript indices:

<table>
<thead>
<tr>
<th>Type of Pair</th>
<th>Number of Pairs</th>
<th>Symbolic Designation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concordant</td>
<td>$\sum_{i,i'} \sum_{i' &gt; i} n_{ii'}$</td>
<td>$P$</td>
</tr>
<tr>
<td>Discordant</td>
<td>$\sum_{i,i'} \sum_{i' &lt; i} n_{ii'}$</td>
<td>$Q$</td>
</tr>
<tr>
<td>Tied on A only</td>
<td>$\sum_{i,i'} n_{ii'}$</td>
<td>$A_T$</td>
</tr>
<tr>
<td>Tied on B only</td>
<td>$\sum_{i,i'} n_{ii'}$</td>
<td>$B_T$</td>
</tr>
<tr>
<td>Tied on A and B (1/2)</td>
<td>$(1/2) \sum_{i,i'} n_{ii'}(n_{ii'} - 1)$</td>
<td>$AB_T$</td>
</tr>
</tbody>
</table>

Using this notation, the symmetric Goodman-Kruskal $\gamma$ can be defined by

$$\gamma = \frac{(P - Q)}{(P + Q)}.$$

In a random selection of two objects from the contingency table, $\gamma$ is equal to the probability of obtaining a concordant pair minus the probability of obtaining a discordant pair, where both probabilities are conditional upon selecting an object pair that is untied. If factor $A$ is treated as an independent factor, and $B$ as a dependent factor, then Somers' asymmetric measure,

$$d_{BA} = \frac{(P - Q)}{(P + Q + B_T)},$$

is again the difference between the two probabilities, but here, the probabilities are conditional upon an object pair being untied solely upon the independent factor. Clearly, if $B$ is treated as the independent factor, then

$$d_{AB} = \frac{(P - Q)}{(P + Q + A_T)}.$$

**FREEMAN'S $\theta$**

If the rows defining factor $A$ are unordered but the columns defining factor $B$ are ordered, Freeman [1965] develops a measure of association $\theta$ by the following argument:

Suppose $i$ and $i'$ are two distinct rows of the contingency table, where $i > i'$; denote the number of concordant and discordant pairs defined solely for these two rows by $P_{ii'}$ and $Q_{ii'}$, respectively, and the absolute value of the excess of concordant over discordant pairs by $|P_{ii'} - Q_{ii'}|$. Thus,

$$P_{ii'} = \sum_{i} \sum_{i' > i} n_{ii'} n_{i'i'};$$

$$Q_{ii'} = \sum_{i} \sum_{i' < i} n_{ii'} n_{i'i'}.$$