The Effect of Category Choice on Some Ordinal Measures of Association

ALAN AGRESTI*

Several ordinal measures of association for cross-classification tables are compared with respect to their stability when various grids are placed on a bivariate normal distribution. Kendall's tau *b* usually fares better than Kendall's tau c, Goodman and Kruskal's gamma, and three extensions of Spearman's rho for cross-classification tables, in terms of approximating an associated measure for ungrouped data. The loss of efficiency of tau *b* due to grouping in testing the hypothesis of no association is considered and observed to be strongly related to the proportion of tied pairs of observations.

1. INTRODUCTION

Several measures have been introduced to describe the association between two variables measured on an ordinal scale (see, e.g., Goodman and Kruskal [3], Kendall [4] and Kruskal [5]), and the pros and cons of using each have been widely debated (see Blalock [1, pp. 421–6] for a partial summary). The numerical values of those measures defined for cross-classification tables depend on the *grid* of the table—the numbers of rows and columns and the choice of categories for each marginal classification. Although the measure should naturally reflect the choice of the grid placed on the given bivariate distribution, it is usually desirable for it to be relatively stable with respect to changes in the nature of the grid if it is to be a reliable index of association. For example, if two researchers choose slightly different categorizations for measuring a pair of variables, hopefully they will reach similar conclusions regarding the strength of the relationship between them.

In this article, we investigate the behavior of some of the most commonly used ordinal measures of association as the grid placed on a bivariate normal distribution with correlation $\rho = .2$, .5 and .8 is varied, by changing

a. the numbers of rows and columns;
b. the marginal categories for each choice of the numbers of rows and columns.

We assume in using this underlying continuous distribution that even if two variables are recorded simply in ordered categories, it often is sensible to interpret the observations on these variables as representing imprecise, underdeveloped, or grouped measurements of interval scale variables, or measurements monotonically related to possibly unobservable interval scale variables.[1] To be reliable in this sense, a measure computed for a cross-classification table should also be similar in value to an associated measure for ungrouped data computed for the underlying continuous distribution.

In addition, the sample size needed to reject the hypothesis of no association at a fixed significance level and power depends on the grid choice. This sample size for a test of no association based on τ_b is calculated for various grids, and a relative efficiency measure is presented by comparing this to a corresponding sample size when the data are ungrouped. The relative efficiency is seen to be approximated by a monotonic function of the proportion of pairs of observations that are untied on both of the rankings. As a special case, the test of no association for a table with a small number of rows or columns is especially inefficient relative to the underlying test for ungrouped data.

2. THE ORDINAL MEASURES AND GRIDS TO BE CONSIDERED

The six measures selected to be compared were those symmetric measures that seem to be most commonly used for describing the strength of the association displayed in a cross-classification table with r ordered row categories and c ordered column categories. Let p_{ij} be the probability that an observation falls in the cell in row i and column j of the table,

$$m = \min(r, c) \ ,$$

$$p_{i.} = \sum_{j=1}^{c} p_{ij} \ , \quad p_{.j} = \sum_{i=1}^{r} p_{ij} \ ,$$

$$F_{k.} = \sum_{i=1}^{k-1} p_{i.} + p_{k.}/2 \ , \quad F_{.k} = \sum_{j=1}^{k-1} p_{.j} + p_{.k}/2 \ ,$$

$$\mu_1 = \sum_{i=1}^{r} i p_{i.} \ , \quad \mu_2 = \sum_{j=1}^{c} j p_{.j} \ ,$$

$$P_c = 2 \sum_{i=1}^{r} \sum_{j=1}^{c} p_{ij} \left(\sum_{i'>i} \sum_{j'>j} p_{i'j'} \right) \ ,$$

$$P_d = 2 \sum_{i=1}^{r} \sum_{j=1}^{c} p_{ij} \left(\sum_{i'>i} \sum_{j'<j} p_{i'j'} \right) \ ,$$

$$P_t = \sum_{i=1}^{r} p_{i.}^2 + \sum_{j=1}^{c} p_{.j}^2 - \sum_{i=1}^{r} \sum_{j=1}^{c} p_{ij}^2 \ .$$

* Alan Agresti is assistant professor, Department of Statistics, University of Florida, Gainesville, Fla. 32611. Computer support was provided by the Northeast Regional Data Center at the University of Florida. The author is grateful to Donald R. Ploch for some helpful comments.

[1] Goodman and Kruskal [3, pp. 735–6] reference some interesting older articles which debate just when such an interpretation is reasonable.

© Journal of the American Statistical Association
March 1976, Volume 71, Number 353
Applications Section

Then, these ordinal measures of association may be defined as follows:

$$\gamma = (P_c - P_d)/(1 - P_t) \ ,$$

$$\tau_b = (P_c - P_d)/((1 - \sum_{i=1}^{r} p_{i.}^2)(1 - \sum_{j=1}^{c} p_{.j}^2))^{\frac{1}{2}} \ ,$$

$$\tau_c = (P_c - P_d)/[(m-1)/m] \ ,$$

$$R = \sum_{i=1}^{r}\sum_{j=1}^{c} (i - \mu_1)(j - \mu_2)p_{ij}/ ((\sum_{i=1}^{r} (i-\mu_1)^2 p_{i.})(\sum_{j=1}^{c} (j-\mu_2)^2 p_{.j}))^{\frac{1}{2}} \ ,$$

$$\rho_b = \sum_{i=1}^{r}\sum_{j=1}^{c} (F_{i.} - .5)(F_{.j} - .5)p_{ij}/ ((\sum_{i=1}^{r} (F_{i.} - .5)^2 p_{i.})(\sum_{j=1}^{c} (F_{.j} - .5)^2 p_{.j}))^{\frac{1}{2}} \ ,$$

$$\rho_c = 1 - 6m^2 \sum_{i=1}^{r}\sum_{j=1}^{c} p_{ij}(F_{i.} - F_{.j})^2/(m^2 - 1) \ .$$

The three measures based on the proportions of concordant and discordant pairs of observations (P_c and P_d) are extensions to cross-classification tables of Kendall's τ_a, which is the difference between these proportions for a continuous bivariate distribution. The proportion of pairs of observations that are tied on at least one of the two rankings $P_t = 1 - (P_c + P_d) > 0$ when the data are grouped, and τ_a uncorrected deflates in value, seriously so when P_t is large. For example, $\tau_a = .333 = \frac{2}{3} - \frac{1}{3}$ in a normal distribution with $\rho = .5$; if each marginal distribution is split at the median, however, the resulting 2×2 table has $P_c = .222$ and $P_d = .056$, so that $\tau_a = .166$; if each marginal distribution is split at the tenth percentile, then $P_c = .055$ and $P_d = .009$, so that $\tau_a = .046$. In the remainder of this article, τ_a denotes $P_c - P_d$ for the underlying normal distribution.

R is the Pearson product moment correlation using integer row and column scores (see [8]) and ρ_b and ρ_c are two extensions of Spearman's rank order correlation coefficient ρ_s to cross-classification tables (see [4, p. 38] and [12]). Notice that ρ_b is the Pearson correlation between the ranks of the two variables using average ranks for the category scores, and thus is the same as R (which treats the row and column numbers as ranks) when, for each variable, the difference between any two adjacent average ranks is the same. τ_c and ρ_c were proposed by Stuart as measures of association which have a maximum absolute value of one for a table of any size, unlike τ_b and ρ_b, which can only attain an absolute value of one when $r = c$.

The table sizes most extensively investigated were $2 \times 2, 2 \times 3, 2 \times 4, 2 \times 5, 2 \times 10, 3 \times 3, 3 \times 4, 3 \times 5, 3 \times 10, 4 \times 4, 4 \times 5, 4 \times 10, 5 \times 5, 5 \times 10$, and 10×10. It was unnecessary to consider tables of size $r > c$, since each measure considered is symmetric in this sense. The row and column categorizations reported in this article were obtained by taking those permutations of probabilities in the following distributions which yield different values for at least one of the six measures.

Categories	*Marginal Probabilities*
2	(.5, .5), (.4, .6), (.3, .7), (.2, .8), (.1, .9)
3	(.333, .333, .333), (.1, .3, .6), (.25, .25, .50)
4	(.25, .25, .25, .25), (.1, .1, .4, .4)
5	.2 each category
10	.1 each category

For example, since each measure has the same value for the 2×3 table with marginal distributions (.4, .6) and (.1, .6, .3) as for the table with marginal distributions (.6, .4) and (.3, .6, .1), one of these was omitted. As a result, 226 distinct grids were considered for each value of ρ in the underlying normal distribution (e.g., 25 2×2 grids, 46 2×3 grids, etc.).[2]

3. STABILITY OF THE MEASURES

In this section we shall observe that as a finer grid is placed on a continuous bivariate distribution, τ_b, τ_c, and γ converge to τ_a, whereas R, ρ_b, and ρ_c converge to ρ_s. We shall consider an ordinal measure of association for a cross-classification table to be stable if, for varied grids, it tends to be close to this limiting value that would be obtained for ordered measurements without ties. When $\rho = .2, .5$, and $.8$, the values of τ_a are .128, .333, and .590 and the values of ρ_s are .191, .483, and .786.

For 2×2 tables, $\tau_b = \rho_b = R = (p_{11}p_{22} - p_{12}p_{21})/(p_{1.}p_{2.}p_{.1}p_{.2})^{\frac{1}{2}}$. This quantity is often denoted by ϕ (see [1, pp. 295–301]), and also equals the square root of the measure τ introduced in [3] for nominal variables. Table 1 illustrates the severe dependence on grid of this measure and of γ, τ_c, and ρ_c for the 2×2 table size.

When both sets of marginal distributions are dichotomized at the median, $p_{11} = \frac{1}{4} + \sin^{-1}(\rho)/2\pi$, and thus, $\tau_b = \tau_c = \rho_b = \rho_c = R$, and these all equal τ_a for the underlying normal distribution. For the grids presented in Table 1, P_t increases as $(.5 - p_{1.})$ increases, $\tau_b(\rho_b, R)$ and τ_c tend to decrease relative to τ_a, whereas γ increases in value above τ_a; ρ_c increases sharply when $\rho = .2$ and $\rho = .5$ and when the marginal distributions are identical when $\rho = .8$, and tends to be far from ρ_s. τ_b is consistently better than τ_c, as would be expected, due to the presence in the denominator of τ_b of a term related to the two marginal distributions. Certainly, though, none of the six measures would be judged to be very stable here as P_t increases.

Table 2 shows the values of the measures of association for various table sizes with marginal row probabilities all equal to $1/r$ and marginal column probabilities all equal to $1/c$. Under these constraints, since

$$P_t = 1/r + 1/c - \sum_{i=1}^{r}\sum_{j=1}^{c} p_{ij}^2 \ ,$$

$$1/\min(r, c) \le P_t \le 1/r + 1/c - 1/rc \ . \quad (3.1)$$

[2] The probabilities in the grids were obtained from [13].

1. Ordinal Measures of Association for Various 2 × 2 Grids on a Bivariate Normal Distribution

ρ	Meas.	Values of $p_{1.}$ in marginal distribution												
		$p_{.1} = p_{1.}$					$p_{.1} = .50$				$p_{.1} = 1 - p_{1.}$			
		.5	.4	.3	.2	.1	.4	.3	.2	.1	.4	.3	.2	.1
.2	γ	.252	.262	.268	.294	.337	.254	.263	.272	.308	.256	.269	.301	.383
	τ_b	.128	.129	.119	.106	.078	.127	.122	.110	.093	.125	.110	.088	.056
	τ_c	.128	.124	.100	.068	.028	.124	.112	.088	.056	.120	.092	.056	.020
	ρ_c	.128	.164	.260	.428	.668	.144	.192	.268	.376	.160	.252	.416	.660
.5	γ	.598	.604	.621	.648	.719	.602	.624	.649	.719	.614	.666	.739	.849
	τ_b	.333	.329	.319	.294	.256	.327	.314	.280	.227	.317	.271	.194	.100
	τ_c	.333	.316	.268	.188	.092	.320	.288	.224	.136	.304	.228	.124	.036
	ρ_c	.333	.356	.428	.548	.732	.340	.368	.404	.456	.344	.388	.484	.676
.8	γ	.877	.875	.881	.902	.917	.879	.905	.936	.968	.896	.945	1.00	1.00
	τ_b	.590	.583	.571	.562	.500	.572	.533	.450	.320	.533	.400	.250	.111
	τ_c	.590	.560	.480	.360	.180	.560	.488	.360	.192	.512	.336	.160	.040
	ρ_c	.590	.600	.640	.720	.820	.580	.568	.540	.512	.552	.496	.520	.680

NOTE: For 2 × 2 tables, $\tau_b = \rho_b = R$.

If the table size is increased so that $r \to \infty$ and $c \to \infty$ and the sequence of grids placed on the continuous bivariate distribution is such that the marginal probabilities have these constraints, then $P_t \to 0$, and P_c and P_d converge to the corresponding values for that continuous distribution, and thus, $\gamma \to \tau_a$, $\tau_b \to \tau_a$, and $\tau_c \to \tau_a$. Under these same conditions, clearly $\rho_b = R$ converges to the Pearson correlation of the two underlying marginal distribution functions $G_1(X)$ and $G_2(Y)$, which is Spearman's ρ_s (grade correlation). Also, ρ_c converges to $1 - 6E[G_1(X) - G_2(Y)]^2$, which is easily seen to be ρ_s. Hence, as $r \to \infty$ and $c \to \infty$ with these row and column marginals, $R \to \rho_s$, $\rho_b \to \rho_s$, and $\rho_c \to \rho_s$. In the case of the bivariate normal density, $\tau_a = (2/\pi) \sin^{-1}(\rho)$ and $\rho_s = (6/\pi) \sin^{-1}(\rho/2)$.

Notice that for the grids summarized in Table 2, τ_b seems to be more stable than τ_c and γ, both in terms of the consistency of the values and closeness to τ_a. Gamma becomes especially inflated for small tables. Also, $\rho_b = R$ tends to be superior to ρ_c when $r \neq c$ in the same two ways. Notice that ρ_c tends to be grossly deflated when m is small (e.g., $2 \times c$ tables). When $r = c$ with these marginals, $\tau_b = \tau_c$ and $\rho_b = \rho_c = R$.

Table 3 summarizes the behavior of the six ordinal measures for the 226 tables described in Section 2. τ_b tends to be closest to the associated measure for ungrouped data, in the sense of having the smallest mean squared error MSE about that value. Notice that MSE increases for each measure as ρ increases. Another way to present the behavior of these measures is to describe the pattern of the values of each measure against P_t. To an approximation, γ is convex increasing in P_t; τ_b and τ_c are concave increasing then decreasing functions of P_t; R and ρ_b are concave decreasing functions of P_t. To a linear approximation, the magnitude of the tendency to increase or decrease is reflected by the slope of the least squares

2. Ordinal Measures of Association for the R × C Grid with $p_{i.} = 1/R$ and $p_{.j} = 1/C$, on a Bivariate Normal Distribution

ρ	Meas.	Grid size $r \times c$												
		2 × 2	2 × 3	2 × 4	2 × 5	2 × 10	3 × 3	3 × 4	3 × 5	4 × 4	4 × 5	5 × 5	5 × 10	10 × 10
.2	γ	.252	.234	.225	.211	.206	.211	.201	.211	.194	.190	.185	.170	.157
	τ_b	.128	.136	.139	.134	.138	.141	.143	.139	.146	.148	.148	.145	.141
	τ_c	.128	.157	.170	.170	.186	.141	.151	.137	.146	.152	.148	.154	.141
	ρ_b	.128	.145	.152	.150	.162	.162	.166	.166	.174	.180	.185	.188	.192
	ρ_c	.128	.065	.045	.030	.026	.162	.143	.189	.174	.170	.185	.176	.192
.5	γ	.598	.559	.537	.517	.487	.527	.507	.487	.478	.467	.450	.419	.391
	τ_b	.333	.344	.347	.339	.332	.365	.370	.364	.368	.370	.366	.360	.354
	τ_c	.333	.398	.424	.429	.446	.365	.392	.399	.368	.382	.366	.381	.354
	ρ_b	.333	.365	.373	.379	.388	.410	.424	.426	.429	.438	.443	.453	.463
	ρ_c	.333	.305	.293	.289	.286	.410	.408	.403	.429	.432	.443	.444	.463
.8	γ	.877	.863	.845	.832	.804	.831	.817	.797	.796	.784	.764	.727	.688
	τ_b	.590	.598	.590	.582	.565	.627	.635	.626	.641	.643	.643	.640	.634
	τ_c	.590	.691	.722	.736	.758	.627	.674	.686	.641	.664	.643	.677	.634
	ρ_b	.590	.635	.646	.650	.660	.686	.707	.709	.726	.737	.742	.751	.767
	ρ_c	.590	.598	.597	.596	.598	.686	.699	.697	.726	.734	.742	.748	.767

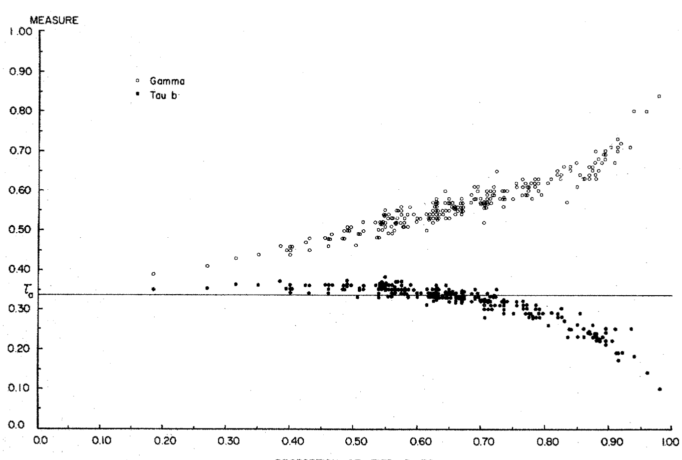
NOTE: For these grids, $\rho_b = R$; when r = c, $\tau_b = \tau_c$, and $\rho_b = \rho_c = R$.

3. Summary of Measure Values and Relationship to Independent Variable P_t, for 226 Grids Placed on a Bivariate Normal Distribution

Statistic	ρ	Measure					
		γ	τ_b	τ_c	R	ρ_b	ρ_c
Mean	.2	.239	.128	.125	.142	.141	.212
	.5	.573	.321	.302	.351	.351	.405
	.8	.869	.538	.506	.578	.582	.620
Standard deviation	.2	.033	.018	.031	.025	.025	.094
	.5	.068	.047	.078	.061	.061	.061
	.8	.053	.097	.142	.116	.116	.076
$(\text{MSE})^{\frac{1}{2}}$ about underlying measure	.2	.116	.018	.032	.056	.056	.096
	.5	.249	.049	.084	.145	.145	.098
	.8	.284	.110	.165	.238	.235	.182
Slope of l.s. line with intercept equal to underlying measure	.2	.166	−.004	−.018	−.077	−.077	−.045
	.5	.364	−.029	−.065	−.207	−.207	−.110
	.8	.432	−.106	−.165	−.346	−.341	−.267

line, which is constrained to equal the measure for ungrouped data when $P_t = 0$. The figure contrasts the values for γ and for τ_b relative to the P_t values, for the underlying $\tau_a = .333$ when $\rho = .5$. This inflationary behavior of γ with respect to collapsings of tables has been noted by many researchers.[3]

Gamma and Tau B [a]



[a] For 226 grids on a normal distribution with $\rho = .5$.

We shall not give a more detailed account of which types of changes in the marginal distributions have the greatest impact on these measures, since this is discussed to some extent by Blalock [2]. His predictions concerning the direction and magnitude of changes in these measures as adjacent categories are combined, are consistent with the results we observed for the bivariate normal distribution.

Another ordinal measure occasionally used for cross-classification tables is ρ_a (see [4, p. 38]), which is equal to $\rho_b((1 - \sum_{i=1}^{r} p_{i.}^3)(1 - \sum_{j=1}^{c} p_{.j}^3))^{\frac{1}{2}}$. Since ρ_b itself typically underestimates ρ_s for the underlying normal distribution, ρ_a would be consistently poorer yet, especially when P_t is large.

If it is necessary to recommend the use of one of these ordinal measures, my choice would be τ_b. It tends to be closer to τ_a for the underlying continuous distribution (at least in the normal case) than γ and τ_c, and closer to τ_a than R, ρ_b, and ρ_c are to ρ_s. In fact, when P_t is large, R, ρ_b, and ρ_c tend to be better approximations for τ_a than for ρ_s. $|\tau_b - \tau_a| \leq .1\tau_a$ for nearly all of the grids considered here for which $P_t \leq .75$. When $P_t > .85$ for a table of any size, one should keep in mind that τ_b could seriously underestimate τ_a, although probably not by as much as γ overestimates it.

Although τ_b does not have as simple an interpretation as γ (the difference in the proportions of concordant and discordant pairs of untied observations), it is meaningful when considered as an approximation for the difference between these two proportions in an underlying continuous distribution, and it can be shown to equal the geometric mean of two useful asymmetric gamma-type measures (see [11]). In addition, τ_b^2 has been given proportional reduction in error interpretations which parallel the one for the square of the Pearson correlation coefficient (see [7, 14]). In fact, τ_b is a natural analog of the Pearson correlation coefficient in a linear model for pairs of observations measured on an ordinal scale, and can be extended naturally to multivariate settings (see, e.g., [7 or 4, Ch. 2 and Ch. 8]).

The criterion of closeness to a corresponding ordinal measure for ungrouped data arises naturally from the assumption of ordinal level measurement. If it is reasonable to assume that there is an underlying higher level of measurement with the bivariate relationship represented by the normal model, one might instead wish to approximate the correlation ρ and consider closeness to it as the criterion of goodness. Since τ_b tends to be closer to τ_a than do τ_c or γ, inversion of the formula $\tau_a = (2/\pi)\sin^{-1}(\rho)$ and substitution of τ_b for τ_a (yielding $\rho \doteqdot \sin(\pi\tau_b/2)$) would usually result in a better approximation for ρ than the corresponding substitution with τ_c or γ. Since

$$\partial\rho/\partial\rho_s = (\partial/\partial\rho_s)[2\sin(\pi\rho_s/6)] \geq (\tfrac{2}{3})(\partial/\partial\tau_a)[\sin(\pi\tau_a/2)] = (\tfrac{2}{3})(\partial\rho/\partial\tau_a) \quad (3.2)$$

whenever $\rho_s \leq 3\tau_a$, this would also produce a better approximation than inverting $\rho_s = (6/\pi)\sin^{-1}(\rho/2)$ and substituting ρ_s' equal R, ρ_b, or ρ_c for ρ_s *at least* when $|\tau_b - \tau_a| \leq \frac{2}{3}|\rho_s' - \rho_s|$ and $\max(\rho_s', \rho_s) \leq 3\min(\tau_b, \tau_a)$. These inequalities hold for ρ_s' equal R, ρ_b, and ρ_c for most of the grids considered of size 3×5 or smaller. Of course, interval level categorical measures (such as the tetrachoric correlation or an analog of R allowing other row and column scores) would also be considered if the observed measurements were themselves of higher level, although easily interpretable ordinal measures such as τ_a

[3] For example, Reynolds [10] has concluded that partial measures of association based on γ are unsatisfactory, due to this tendency to overstate the true relationship. Quade [9] shows that an explanation of these higher values is the fact that γ treats $P_c/(P_c + P_d)$ of the tied pairs as concordant and $P_d/(P_c + P_d)$ as discordant. Ordinarily, one would not expect a pair of observations from a subpopulation in which at least one of the variables is restricted in range to exhibit as strong an association as a pair of observations picked at random from the entire population.

are even then descriptively very useful for nonnormal relationships.

4. EFFICIENCIES FOR CROSS-CLASSIFICATION TABLES

For a given grid, the random sample version of each measure is asymptotically normally distributed about the population value with variance depending on the grid and underlying distribution and inversely proportional to sample size. Proctor compared the sample size required for each of a collection of ordinal measures to attain equal power in rejecting the null hypothesis of independence of row and column categories for some tables based primarily on an underlying normal distribution with $\rho = .8$ and a model for measurement error. In comparing γ, τ_b, τ_c, and R, he concluded that

> Generally speaking, under certain conditions one measure will be best while under other conditions another will be .... While variations in the error process and in the underlying pattern of association will lead the relative efficiencies to change a bit, the four measures of association are all quite similar [8, p. 378].

The introduction of measurement error had some effect in improving the efficiency of an ordinal measure such as γ relative to R, which is based on equal intervals between category scores.

Naturally, the efficiency of each ordinal measure depends on the grid. We investigated the nature of the change in the asymptotic sampling distribution of t_b (the random sample version of τ_b), using the grids and underlying normal distributions of Section 3. The asymptotic variance[4] of t_b is σ^2/n (see [8]), where

$$\begin{aligned}\sigma^2 = {} & (4/\Delta_1\Delta_2)\{(P_{cc} - 2P_{cd} + P_{dd}) \\ & - (P_c - P_d)^2(1/\Delta_1 + 1/\Delta_2)^2/4 \\ & + (P_c - P_d)\sum_{i=1}^{r}\sum_{j=1}^{c} p_{ij}(\sum_{i'>i}\sum_{j'>j} p_{i'j'} + \sum_{i'<i}\sum_{j'<j} p_{i'j'} \\ & - \sum_{i'>i}\sum_{j'<j} p_{i'j'} - \sum_{i'<i}\sum_{j'>j} p_{i'j'})(p_{i.}/\Delta_1 + p_{.j}/\Delta_2) \\ & + (P_c - P_d)^2 \sum_{i=1}^{r}\sum_{j=1}^{c} p_{ij}(p_{i.}/\Delta_1 + p_{.j}/\Delta_2)^2/4\} \ , \end{aligned} \tag{4.1}$$

and

$$P_{cc} = \sum_{i=1}^{r}\sum_{j=1}^{c} p_{ij}(\sum_{i'>i}\sum_{j'>j} p_{i'j'} + \sum_{i'<i}\sum_{j'<j} p_{i'j'})^2 \ ,$$

$$P_{dd} = \sum_{i=1}^{r}\sum_{j=1}^{c} p_{ij}(\sum_{i'>i}\sum_{j'<j} p_{i'j'} + \sum_{i'<i}\sum_{j'>j} p_{i'j'})^2 \ ,$$

$$\begin{aligned}P_{cd} = {} & \sum_{i=1}^{r}\sum_{j=1}^{c} p_{ij}(\sum_{i'>i}\sum_{j'>j} p_{i'j'} + \sum_{i'<i}\sum_{j'<j} p_{i'j'}) \\ & \cdot(\sum_{i'>i}\sum_{j'<j} p_{i'j'} + \sum_{i'<i}\sum_{j'>j} p_{i'j'}) \ , \end{aligned}$$

$$\Delta_1 = 1 - \sum_{i=1}^{r} p_{i.}^2 \ , \quad \Delta_2 = 1 - \sum_{j=1}^{c} p_{.j}^2 \ .$$

[4] The formula as presented by Proctor [8] is printed incorrectly (part of it is missing), but it does not seem to be presented anywhere else in the literature. A derivation is available from this author.

The sample size needed to attain a fixed power at a fixed significance level for the null hypothesis of no association ($\tau_b = 0$) is then approximately

$$n = c\sigma^2/\tau_b^2 \ , \tag{4.2}$$

where c is a constant related to these levels.

We compared this sample size to the standard of the sample size needed for the same test based on ungrouped measurements from a normal population. Then the asymptotic variance of t_a (the sample version of $\tau_a = \tau_b$) in this continuous case is σ_0^2/n (Kendall [4, p. 126]), where

$$\sigma_0^2 = 4\{\tfrac{1}{9} - [(2/\pi)\sin^{-1}(\rho/2)]^2\} \ . \tag{4.3}$$

The sample size required here to achieve the same power at the same significance level as previously is approximately

$$n_0 = c\sigma_0^2/\tau_a^2 \ . \tag{4.4}$$

The asymptotic efficiency of the test based on grouped data relative to the test based on ungrouped data can then be defined by the ratio

$$\text{RE} = n_0/n = \sigma_0^2\tau_b^2/\sigma^2\tau_a^2 \ . \tag{4.5}$$

A comparison of RE values for various grids gives insight into one of the types of information loss that occurs in grouping data or collapsing categories.

Table 4 presents the relative efficiencies for the 25 2×2 tables, when $\rho = .5$. Even for the best 2×2 table (when both marginal distributions are split at the median), the grouped data procedure requires $1/.380 = 2.63$ times as many observations as the ungrouped data procedure. The situation deteriorates as the cutting point for each marginal distribution is drawn away from the median; when $p_{1.} = p_{.1} = .10$, for example, RE $= .095$.

Results similar to those in Table 4 occur when $\rho = .2$ and $\rho = .8$, with the test for grouped data performing poorest relative to the test for ungrouped data when

4. Asymptotic Efficiencies for a Test of No Association ($\tau_b = 0$) in a 2×2 Table Relative to the Test for Ungrouped Data, Based on a Normal Distribution with $\rho = .5$

$p_{1.}$	$p_{.1}$ above diagonal: .50	.40	.30	.20	.10
.50	.380	.372	.359	.305	.234
.40	.376	.360	.334	.285	.208
.30	.356	.360	.301	.247	.171
.20	.322	.334	.294	.202	.134
.10	.226	.297	.231	.182	.095
$p_{.1}$ below diagonal	.60	.70	.80	.90	

$\rho = .8$ and best when $\rho = .2$. For example, when $\rho = .2$, RE = .434 for the 2×2 table with $p_{1.} = p_{.1} = .50$, and RE = .106 for the 2×2 table with $p_{1.} = p_{.1} = .10$; when $\rho = .8$, the corresponding values are .263 and .059. For all values of ρ, as r and c increase in such a way that P_t decreases toward zero, RE increases toward one. For example, when $\rho = .5$ with $p_{i.} = 1/r$ and $p_{.j} = 1/c$, the test for the 4×4 table is about twice as efficient as the test for the 2×2 table (.778 vs. .380), and RE = .926 for the 10×10 table.

A more thorough inspection of the 226 grids for each value of ρ reveals that the RE values are linearly related to the $1 - P_t$ values to a good approximation (see Table 5). For example, the Pearson correlation between RE and $1 - P_t$ equals .956 when $\rho = .5$. Now to the extent that the number of untied pairs of observations is a measure of the available information for a test of no association, one might expect that n_0 observations (which yield $n_0(n_0 - 1)/2$ pairs) in the test for ungrouped data are equivalent to n observations in the test for grouped data, where

$$n_0(n_0 - 1)/2 \doteq (1 - P_t)n(n - 1)/2 \ .$$

This implies that RE should be on the order of $(1 - P_t)^{\frac{1}{2}}$. Scatter diagrams between $1 - P_t$ and RE for τ_b display slight concave deviations from linearity when $\rho = .2$ and $\rho = .5$, and in fact, the root mean square error of RE about $(1 - P_t)^{\frac{1}{2}}$ is not much larger than about the least squares line for these cases.

5. Relationship Between RE and $1 - P_t$ for τ_b, Calculated for 226 Grids Placed on a Bivariate Normal Distribution with Correlation ρ

ρ	Pearson $r_{RE,1-P_t}$	Least squares line	$(MSE)^{\frac{1}{2}}$ of RE about	
			l. s. line	$(1 - P_t)^{\frac{1}{2}}$
.2	.973	$.15 + 1.26(1 - P_t)$	.042	.060
.5	.956	$.13 + 1.16(1 - P_t)$	.050	.081
.8	.938	$-.03 + 1.31(1 - P_t)$	.070	.179

Since γ and τ_c are similar in structure to τ_b and approximately equal in efficiency according to Proctor [8], the results in this section can also be interpreted as an indication of the dependence of the efficiency for these measures on the grid. In particular, one could conjecture that $1 - (1 - P_t)^{\frac{1}{2}}$ is a crude but simple measure for approximating the relative loss of efficiency (1 − RE) due to grouping for such ordinal measures in testing the hypothesis of no association, at least when the observations are taken from a bivariate normal distribution with small to moderate correlation.

5. CONCLUSIONS

Many criteria have been applied by social scientists and statisticans to evaluate the various measures of association, including the following.

i. meaningful operational interpretation (such as proportional reduction in error),
ii. simplicity of the measure and its sampling distribution,
iii. sensitivity to form of distribution,
iv. efficiency in rejecting the null hypothesis of no association.

A criterion that unfortunately seems to be given greater importance by many practitioners is size of the measure—the one with the largest absolute value being chosen because it makes the results seem more striking.

This article emphasizes that in many situations an important criterion is that a measure computed for a cross-classification table should be similar in value to a corresponding measure that would be computed if the data were measured more precisely. Thus, the measure should be reasonably stable as various grids are placed on the underlying distribution. According to this criterion, perhaps the most commonly used ordinal measure (gamma) fares very poorly for a wide range of tables based on a bivariate normal distribution. Although tables can be artificially constructed so that any one of these measures appears better than the others, and the results in this article relate strictly only to those grids and those three normal distributions considered, these results are consistent with other more limited investigations of this behavior (see [2, 9, 10]). A natural extension of this study would be the consideration of grouping effects for a less structured underlying distribution, perhaps one of the families listed in [6].

In summary, Kendall's τ_b seems more stable overall than the others in terms of approximating the corresponding measure for ungrouped data. Also, whenever possible, care should be exercised in choosing a grid for a table. When the number of rows or columns is increased or the marginal proportions are selected to minimize the proportion of tied pairs of observations, the efficiency of a hypothesis test of no association tends to improve and the value of the measure tends to be closer to the value for the underlying continuous distribution.

[*Received December 1974. Revised June 1975.*]

REFERENCES

[1] Blalock, Hubert M., *Social Statistics*, Second ed., New York: McGraw-Hill Book Co., 1972.
[2] ———, "Beyond Ordinal Measurement: Weak Tests of Stronger Theories," in Hubert M. Blalock, ed., *Measurement in the Social Sciences*, Chicago: Aldine-Atherton, 1974, Ch. 15.
[3] Goodman, Leo A. and Kruskal, William H., "Measures of Association for Cross-Classification," *Journal of the American Statistical Association*, 49 (December 1954), 732–64.
[4] Kendall, Maurice G., *Rank Correlation Methods*, Fourth ed., London: Charles W. Griffin & Co., Ltd., 1970.
[5] Kruskal, William H., "Ordinal Measures of Association," *Journal of the American Statistical Association*, 53 (December 1958), 814–61.
[6] Plackett, R.L., "A Class of Bivariate Distributions," *Journal of the American Statistical Association*, 60 (June 1965), 516–22.
[7] Ploch, Donald R., "Ordinal Measures of Association and the General Linear Model," in Hubert M. Blalock, ed., *Measurement in the Social Sciences*, Chicago: Aldine-Atherton, 1974, Ch. 12.

[8] Proctor, Charles H., "Relative Efficiencies of Some Measures of Association for Ordered Two-Way Contingency Tables Under Varying Intervalness of Measurement Errors," *Proceedings of the Social Statistics Section, American Statistical Association* (1973), 372–9.

[9] Quade, Dana, "Nonparametric Partial Correlation," in Hubert M. Blalock, ed., *Measurement in the Social Sciences*, Chicago: Aldine-Atherton, 1974, Ch. 13.

[10] Reynolds, H.T., "Ordinal Partial Correlation and Causal Inferences," in Hubert M. Blalock, ed., *Measurement in the Social Sciences*, Chicago: Aldine-Atherton, 1974, Ch. 14.

[11] Somers, Robert H., "A Similarity Between Goodman and Kruskal's Tau and Kendall's Tau, with a Partial Interpretation of the Latter," *Journal of the American Statistical Association*, 57 (December 1962), 804–12.

[12] Stuart, Alan, "Calculation of Spearman's Rho for Ordered Two-Way Classifications," *American Statistician*, 17 (October 1963), 23–4.

[13] U.S. National Bureau of Standards, *Tables of the Bivariate Normal Distribution Function and Related Functions*, Washington, D.C.: U.S. Government Printing Office, 1959.

[14] Wilson, Thomas P., "A Proportional Reduction in Error Interpretation for Kendall's Tau-*b*," *Social Forces*, 47 (March 1969), 340–2.