RELATIVE EFFICIENCIES OF SOME MEASURES OF ASSOCIATION FOR ORDERED TWO-WAY CONTINGENCY TABLES UNDER VARYING INTERVALNESS OF MEASUREMENT ERRORS

Charles H. Proctor, North Carolina State University

1. Introduction

As one who fairly often sees contingency tables in connection with analyzing social survey data, I recently became interested in what computations to include in the local statistical package to accompany the cross tabulation output. This paper reports on the approach that was developed to evaluate measures of association for two-way tables with ordered variates in both directions. Generally speaking, under certain conditions one measure will be best, while under other conditions another one will be. This is not an unexpected conclusion, but what may be new here is the emphasis on a somewhat more empirical method than any I have seen heretofore, in describing these conditions.

For the case of continuous variates there are results already available on the relative efficiencies of several measures of association. One major problem of such work has been to characterize, realistically and parsimoniously, the case of association or of non-independence. It has been handled by Konijn [14] and Farlie [7] in somewhat different ways. For contingency tables, the distributions suggested by Plackett [17] and referred to by Mosteller [19] would seem a natural basis for investigating power, but they were not used here. I am not aware that they have been used to exhibit relative efficiencies.

Actually, it seems to be that two approaches are being followed in the selection of a measure of association. One, employed by Kruskal [15], emphasizes that the measure computed on sample frequencies estimates a counterpart population quantity and the user should be sure he wants to know that population quantity. Interpretations are provided of these population quantities in contexts such as predicting the ordering of a pair of persons on a second IQ test from their ordering on a first IQ test. The underlying method involves discovering what actions the user wants to take when he sees his data, and then verifying that the suggested statistic will fit into that action pattern. Although the admonition to take into account the user's interests is undeniably good, the method still has its ambiguities. The amount of controversy generated, by the choice of a measure of association, among sociological methodologists is, I believe, testimony to these ambiguities [6, 16], although this issue is properly a problem in the sociology of knowledge.

Another approach is by way of measurement model theory. The investigator states what he judges to be the measurement scale of his variates, and then if, for example, both the row and column variates are of ordinal types, but not interval, he calculates the Goodman-Kruskal gamma or the Kendall tau-sub-b. At any rate, a Pearson product moment coefficient would be meaningless (a technical term [21, p. 66]) unless his scales were interval. The definitions of such distinctions among variables are most elegantly expressed as equivalence classes under certain transformations [21, p. 10]. Thus, the scale is of ordinal type if it is equivalent in distinguishing among observational units, to any other scale obtained by a monotonic transformation of it. By substituting affine transformation in place of monotonic transformation in the above, one defines an interval scale type.

The present work advocates the philosophy of this second approach, but suggests using empirical evidence in the data themselves for characterizing the measurement model type. The absence of these empirical criteria for determining scale type has always struck me as a shortcoming of the approach, and the works of Suppes and Zinnes [21, pp. 72-74] and Campbell [5, Chapters XVII and XVIII] in their final discussions of random measurement errors, serve as the stepping-off-place for the present development. The data are here viewed as sampled from a population table in which the cell probabilities are formed in part by a parent, generic stochastic process and in another part by random drift among cells arising from measurement error (also called misclassification error by Mote and Anderson [18] in this categorical variable context). The criterion to be employed for judging the measures is relative efficiency and these efficiencies are to be calculated for variations in the population table.

2. Defining the Parent Process, the Error Process and the Sampling Process

The starting point for the comparison of measures of association is a two-way contingency table, an A by B table, whose row and column variates are categorical, but ordered. The statuses of the two qualitative variables are taken to be causally symmetric, that is, jointly dependent, either on one another or on some collection of unmeasured independent variables. The underlying process of interdependence will be supposed to have produced a joint distribution of the units of observation with cell probabilities π_{ij}, with $\sum_{i=1}^{A} \sum_{j=1}^{B} \pi_{ij} = 1$. The probability π_{ij} reflects the chance that an observational unit will emerge with the i^{th} category of the row variate and the j^{th} category of the column variate. The π_{ij} will be taken to reflect the parent stochastic process.

Now further suppose that the measurement operation is to some extent fallible or that, after emergence, the unit's characteristics drift in a random fashion. This process will be called the error process. In particular, the chance of recording row category a when the unit is actually in category i will be denoted by θ_{ia}. Similarly Φ_{jb} is taken to be the chance that a unit that emerged in column category j is recorded as being in category b. Combining the two processes by supposing the errors to be independent gives the resulting probability of observing a unit in cell (a,b) as:

$$(2.1) \qquad \rho_{ab} = \sum_i \sum_j \theta_{ia}\Phi_{jb}\pi_{ij} \ .$$

If the error processes are not independent equation (2.1) becomes:

$$(2.2) \qquad \rho_{ab} = \sum_i \sum_j \delta_{ij,ab}\ \pi_{ij} \ ,$$

which may look simpler but involves many more misclassification parameters in the $\delta_{ij,ab}$'s than in the θ_{ia}'s and Φ_{jb}'s. Such a model was used by Assakul and Proctor [2] to show how great a loss of power the chi-square test suffers under measurement error. It is also presented by Hayashi [12] who corrects biases in cross tabulations by estimating the misclassification parameters.

Both notions, the one of errors and the other of the joint parent distribution, need more empirical content if they are to be usable by the data handler. Measurement errors can be examined by duplicating measurements either by using a superior or optimal measurement method or by two parallel applications of the usual technique. From the sets of duplicate determinations one can estimate the θ_{ia} and the Φ_{jb}. Discussion of this estimation problem, important as it is, would lead us too far afield. Thus we will merely suggest the model equations for measurement and drift errors that will be used to define intervalness and leave as a separate task, the estimation of the parameters in such a model.

It is taken that the misclassification probabilities follow the pattern of:

$$\theta_{ia} = (A-1)\alpha_1 e^{-\beta_1|i-a|} \Big/ \Big(\sum_{i\neq a} e^{-\beta_1|i-a|}\Big), \text{ if } i \neq a$$

$$(2.3)$$

$$= 1 - (A-1)\ \alpha_1 \quad \text{when } i = a \ .$$

This may look complex but it merely states that θ_{ia} depends firstly on an overall level of error, measured by α_1. If β_1 is large this error occurs largely in adjacent categories, while if β_1 is small then distant categories may be confused. Thus, if α_1 is zero, there is no measurement nor drift error, while if β_1 is zero the pattern is of "ordinal" type, and the larger β_1 becomes the more the pattern can be called "interval".

To complete the picture the Φ_{jb}'s may be similarly defined by:

$$\Phi_{jb} = (B-1)\alpha_2 e^{-\beta_2|j-b|} \Big/ \Big(\sum_{j=b} e^{-\beta_2|j-b|}\Big), \text{ if } j \neq b$$

$$(2.4)$$

$$= 1 - (B-1)\ \alpha_2 \quad \text{when } j = b \ .$$

However, in all of the following calculations we have taken $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$.

Having introduced the generic or parent process represented by π_{ij} and the drift or measurement error process by the θ_{ia}'s and Φ_{jb}'s, it remains to specify the sample selection process. If one supposes that each of the n observational units had the same chance, namely π_{ij}, of being in the (i,j) cell, and was subject to the same error process and that these processes acted independently from one unit to the next, then the cell frequencies, the n_{ab}, would be distributed in accord with a multinomial distribution with AB classes having underlying probabilities $\{\rho_{ab}\}$. And this is what we will assume. Of the three sampling distributions cases distinguished by Barnard [4] as (a) both margins fixed, (b) one fixed, and (c) both free, ours is the third.

3. Measures of Association to be Compared

The four principal measures of association to be examined are the Goodman-Kruskal G, Kendall's TB, the Kendall-Stuart TC, and a coefficient R that is a Pearson product-moment correlation coefficient using integer row and column scores. The first three appear to be widely used in the social sciences, while the fourth is congenial to the author's naive numerical point of view and follows Williams' [22] and Yates' [23] approaches. The definitional formulas are (see [11, p. 325; 10, p. 751; and 13, p. 563]):

$$(3.1) \quad G = (P_s - P_d)/(P_s + P_d)$$

(3.2) $TC = (P_s - P_d)[A/(A - 1)]$

(3.3) $TB = (P_s - P_d)/[(1 - \Sigma R^2_{a.})(1 - \Sigma R^2_{.b})]^{1/2}$

(3.4) $R = \sum_{a=1}^{A} \sum_{b=1}^{B} (a - \bar{a})(b - \bar{b}) R_{ab}/s_a s_b$,

where the quantities involved in these formulas are defined as:

$$P_s = 2 \sum_a \sum_b R_{ab}\{ \sum_{a'>a} \sum_{b'>b} R_{a'b'}\} ,$$

$$P_d = 2 \sum_a \sum_b R_{ab}\{ \sum_{a'>a} \sum_{b'<b} R_{a'b'}\} ,$$

$$R_{a.} = \sum_{b=1}^{B} R_{ab} \quad \text{and} \quad R_{.b} = \sum_{a=1}^{A} R_{ab} ,$$

with

$$s_a^2 = \sum_{a=1}^{A} (a - \bar{a})^2 R_{a.}$$

and

$$s_b^2 = \sum_{b=1}^{B} (b - \bar{b})^2 R_{.b} ,$$

where

$$\bar{a} = \sum_a a\, R_{a.} \quad \text{and} \quad \bar{b} = \sum_b b\, R_{.b} .$$

4. Basis of Comparison, Relative Efficiency

The choice of criteria for comparing one measure with the other is not so much a mathematical question nor an empirical one, it is more like a moral one. In accord with cannons of argument from tradition in this ethical field, I will adduce that the criterion, namely efficiency, that I will use is widely accepted in the statistical profession. In comparing two procedures one frequently computes the ratio of the two sample sizes required to attain the same variance in estimating some parameter or required to achieve the same power in testing some hypothesis. This ratio is called relative efficiency and since larger samples are generally more costly, it shows which procedure will, in that sense, make best use of the observations. In the present case it is somewhat difficult to decide whether the problem is one of estimating a parameter or of testing an hypothesis.

As defined, the four measures are statistics; they are functions of the sample relative frequencies, the R_{ab}'s. In each case there is the same function of the population relative frequencies, the ρ_{ab}'s, that constitutes a population parameter. However, there are four different parameters, and it would seem necessary first to decide which parameter was needed (as Kruskal [16] does) and then use the corresponding statistic.

The alternative point of view we have adopted is that the null hypothesis, H_0 is, in all cases, that of the independence of row and column categories, while the investigator is interested in detecting departures from H_0. From his knowledge, experience and perhaps a glance at the data, he suggests a structure on the π_{ij} and supposes some pattern of θ_{ia} and Φ_{jb}. We thus arrive at a set of ρ_{ab}, the alternative hypothesized population values. The sample is assumed drawn from this population and the measure is computed. It is proposed to compare any two such measures by the ratio of sample sizes that would be required to attain the same power at the alternative hypothesized values.

The reasoning goes that each statistic, being a function of cell relative frequencies, is asymptotically, as sample size increases, normally distributed (about zero if H_0 holds) and one would divide the statistic by its standard deviation to obtain a critical ratio. He would then refer this Z-value, say, to a table of areas under the normal curve to furnish a significance probability for rejecting H_0. If the test is made at a 5% level of significance then a distance of Z = 1.960 (where $\Phi(1.960) = .975$ and Φ is the normal distribution function) away from zero would lead to rejecting H_0. In order to assure this rejection with fairly high probability of, say, .80 would require the population mean Z-value to be 1.960 plus .842 (since $\Phi(.842) = .80$), or 2.802 away from zero. The population mean Z-value is the parameter value under the alternative hypothesis divided by the standard deviation. In all cases, the standard deviations of the statistics, based as they (the statistics) are upon sample proportions, are, to a first approximation, proportional to $n^{-1/2}$. Thus the required sample sizes become (note that $2.802^2 = 7.85$):

(4.1) $$n_G = 7.85\, V_G/\gamma^2, \quad n_{TC} = 7.85\, V_{TC}/\tau_c^2 ,$$
$$n_{TB} = 7.85\, V_{TB}/\tau_b^2, \quad n_R = 7.85\, V_R/\rho^2$$

where the variances are given as $V(G) = V_G/n$, $V(TC) = V_{TC}/n$, etc. and γ^2, τ_c^2, τ_b^2 and ρ^2 are given by (4.1) to (4.4) with ρ_{ab} in place of R_{ab}. When comparing these sample sizes, one sees that ratios such as γ^2/V_G, τ_c^2/V_{TC}, etc. are the crucial quantities. These ratios will be denoted as the "precisions" of the statistics with notation:

(4.2) $$P_G = \gamma^2/V_G, \quad P_c = \tau_c^2/V_{TC},$$
$$P_B = \tau_b^2/V_{TB}, \quad \text{and} \quad P_R = \rho^2/V_R .$$

The argument here is closely akin to that of Bahadur [3] in which his quantity c, or "slope", is equal to γ^2/V_G or τ_c^2/V_{TC}, and so forth.

5. Variance Expressions and Derivations

In the case of G, Goodman and Kruskal [11] give

$$(5.1)\quad V_G = nV(G) = 16[\Pi_s^2\Pi_{dd} - 2\Pi_s\Pi_{sd} + \Pi_d^2\Pi_{ss}]/(\Pi_s + \Pi_d)^4 ,$$

where

$$\Pi_{ss} = \sum_a\sum_b \rho_{ab}[\sum_{a'>a}\sum_{b'>b}\rho_{a'b'} + \sum_{a'<a}\sum_{b'<b}\rho_{a'b'}]^2 ,$$

$$\Pi_{dd} = \sum_a\sum_b \rho_{ab}[\sum_{a'>a}\sum_{b'<b}\rho_{a'b'} + \sum_{a'<a}\sum_{b'>b}\rho_{a'b'}]^2 ,$$

and

$$\Pi_{sd} = \sum_a\sum_b \rho_{ab}[\sum_{a'>a}\sum_{b'>b}\rho_{a'b'} + \sum_{a'<a}\sum_{b'<b}\rho_{a'b'}] \times [\sum_{a'>a}\sum_{b'<b}\rho_{a'b'} + \sum_{a'<a}\sum_{b'>b}\rho_{a'b'}] ,$$

and where Π_s and Π_d are the counterparts of P_s and P_d in formulas (4.5) and (4.6) when ρ_{ab} is used in place of R_{ab}. Formula (5.1) can be derived by a straight forward, if tedious, application of a method described by R. A. Fisher [8, p. 309-310] for the variance of any statistic, T say, when T is a function of frequencies that obey a multinomial distribution. Fisher's formula is:

$$(5.2)\quad \frac{1}{n}V(T) = \sum_a\sum_b\{N_{ab}(\frac{\partial T}{\partial N_{ab}})^2\} - [\frac{\partial T}{\partial n}]^2 \Big|_{N_{ab} = n\rho_{ab}} .$$

The last term is zero for three of the four statistic since n enters explicitly only into TB. For the other three, the formula can be written in terms of relative frequencies as:

$$(5.3)\qquad nV(T) = \sum\sum R_{ab}(\frac{\partial T}{\partial R_{ab}})^2 .$$

The results are:

$$(5.4)^1\ V_{TC} = (\frac{A}{A-1})^2 4[\Pi_{ss} - 2\Pi_{sd} + \Pi_{dd} - (\Pi_s - \Pi_d)^2] .$$

After quite messy algebra it turns out that:

$$(5.5)\quad V_{TB} = \frac{4}{\Delta_a\Delta_b}\{(\Pi_{ss} - 2\Pi_{sd} + \Pi_{dd}) + (\Pi_s - \Pi_d) \times \sum_a\sum_b \rho_{ab}(\sum_{(ab)}^{(s)}\rho_{a'b'})(\frac{\rho_{.b}}{\Delta_b} + \frac{\rho_{a.}}{\Delta_a}) + \frac{1}{4}(\Pi_s - \Pi_d)^2\sum_a\sum_b\rho_{ab}(\frac{\rho_{.b}}{\Delta_b} + \frac{\rho_{a.}}{\Delta_a})^2\} - \frac{(\Pi_s - \Pi_d)^2}{\Delta_a\Delta_b}[\frac{1}{\Delta_b} + \frac{1}{\Delta_a}]^2 ,$$

where $\Delta_a = 1 - \Sigma\rho_{a.}^{\ 2}$ and $\Delta_b = 1 - \Sigma\rho_{.b}^{\ 2}$, while the result for R is:

$$(5.6)\quad V_R = \{(1 + \rho^2/2)\sum_a\sum_b(a-\bar{a})^2(b-\bar{b})^2\rho_{ab} - \rho\sigma_a\sum\sum(a-\bar{a})(b-\bar{b})^3\rho_{ab}/\sigma_b - \rho\sigma_b\sum\sum(a-\bar{a})^3(b-\bar{b})\rho_{ab}/\sigma_a + \frac{\rho^2}{4}(\sigma_a^2\sum_b(b-\bar{b})^4\rho_{.b}/\sigma_b^2 + \sigma_b^2\sum_a(a-\bar{a})^4 \times \rho_{a.}/\sigma_a^2)\}/\sigma_a^2\ \sigma_b^2 .$$

A comment may be in order on the status, as approximations, of the quantities V_G, V_{TC}, V_{TB} and V_R. The random variables G, TC, TB and R have distributions induced, as mentioned before, by the multinomial distribution of the R_{ab}. Their variances are complicated functions of the ρ_{ab}; in fact, except for TC, the expressions are infinite series, the terms of which can be collected in increasing powers of n^{-1}. The expressions for V_G, V_{TC}, V_{TB}, and V_R are the coefficients of n^{-1} in these series. As n increases the other terms become smaller at a rate faster than this first term. Basing our comparison on this first term implies that our results hold only for large sample sizes.[2]

6. Relative Efficiency of the Chi-Square Contingency Table Test

In order to complete the picture, the chi-square test statistic for contingency tables has been included in the comparisons of efficiencies. The statistic is:

$$(6.1)\quad X^2 = n\sum_a\sum_b(R_{ab} - R_{a.}R_{.b})^2/R_{a.}R_{.b} .$$

The, so called, non-centrality parameter is:

$$(6.2)\qquad \lambda = \sum\sum(\rho_{ab} - \rho_{a.}\rho_{.b})^2/\rho_{a.}\rho_{b.}$$

By an argument similar to that which led to formulas (4.1) only based on the non-central chi-square distribution, the formula for required sample size is:

$$(6.3)\qquad n_{X^2} = \lambda(DF, .05, .80)/\lambda .$$

Here λ(DF, .05, .80) is an entry from tables by E. Fix [9], which is the value of the non-centrality parameter for the non-null distribution of X^2 such that a test at $\alpha = .05$ based on X^2 will reject H_0 with probability .80. In particular, λ(25, .05, .80) = 22.843 is the value we will use when discussing some 6 by 6

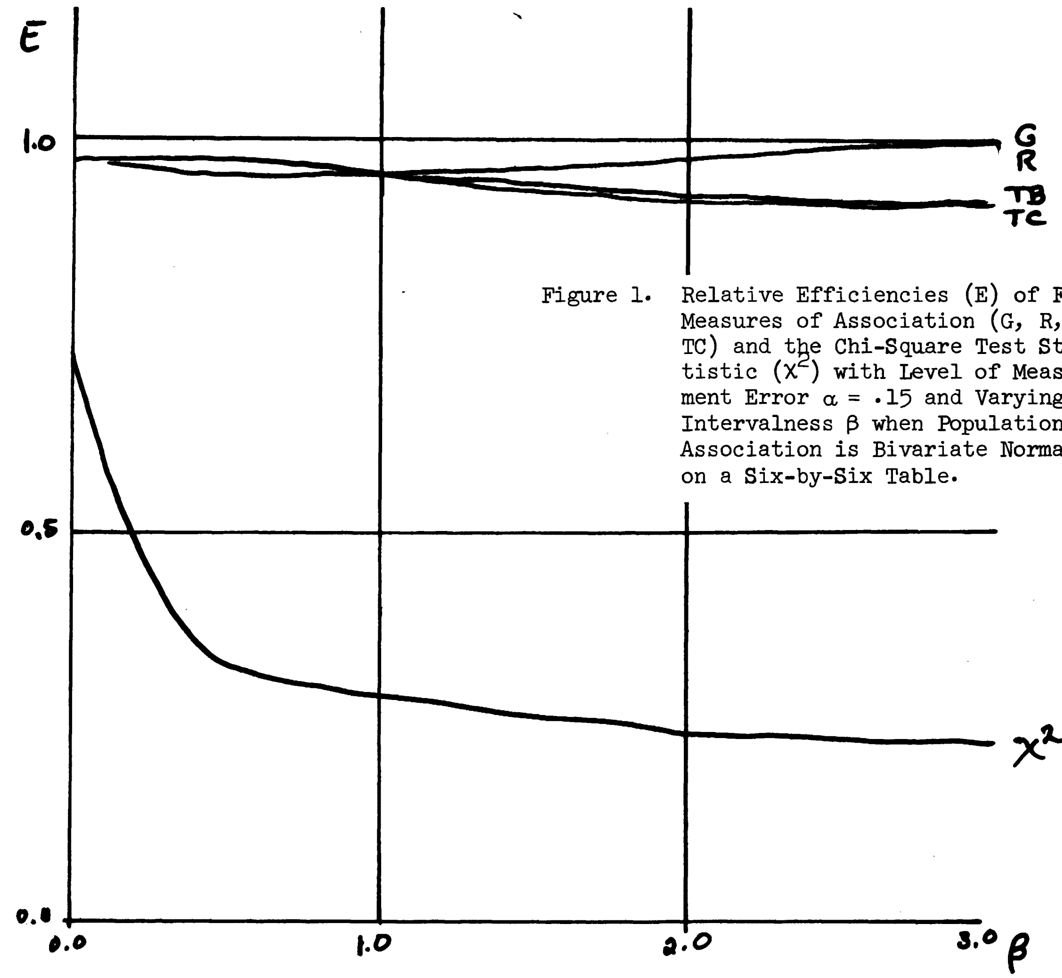

Figure 1. Relative Efficiencies (E) of Four Measures of Association (G, R, TB, TC) and the Chi-Square Test Statistic (χ^2) with Level of Measurement Error $\alpha = .15$ and Varying Intervalness β when Population Association is Bivariate Normal on a Six-by-Six Table.

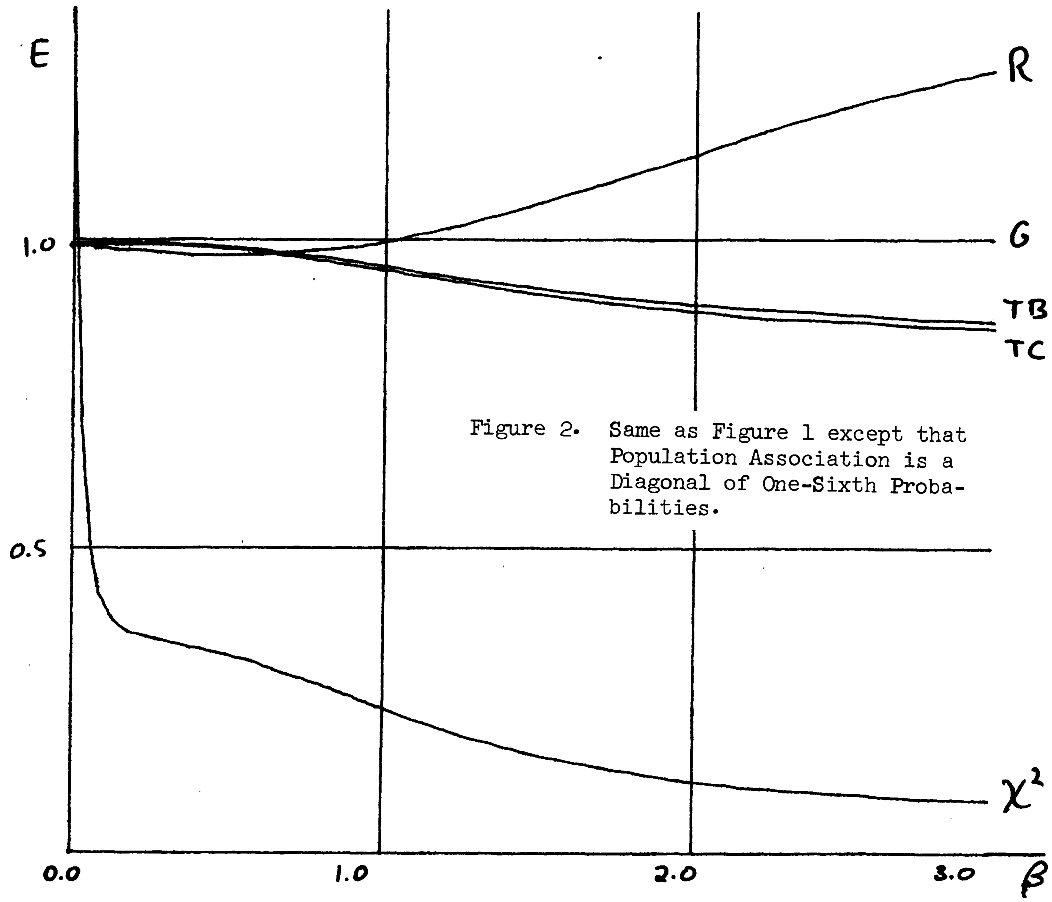

Figure 2. Same as Figure 1 except that Population Association is a Diagonal of One-Sixth Probabilities.

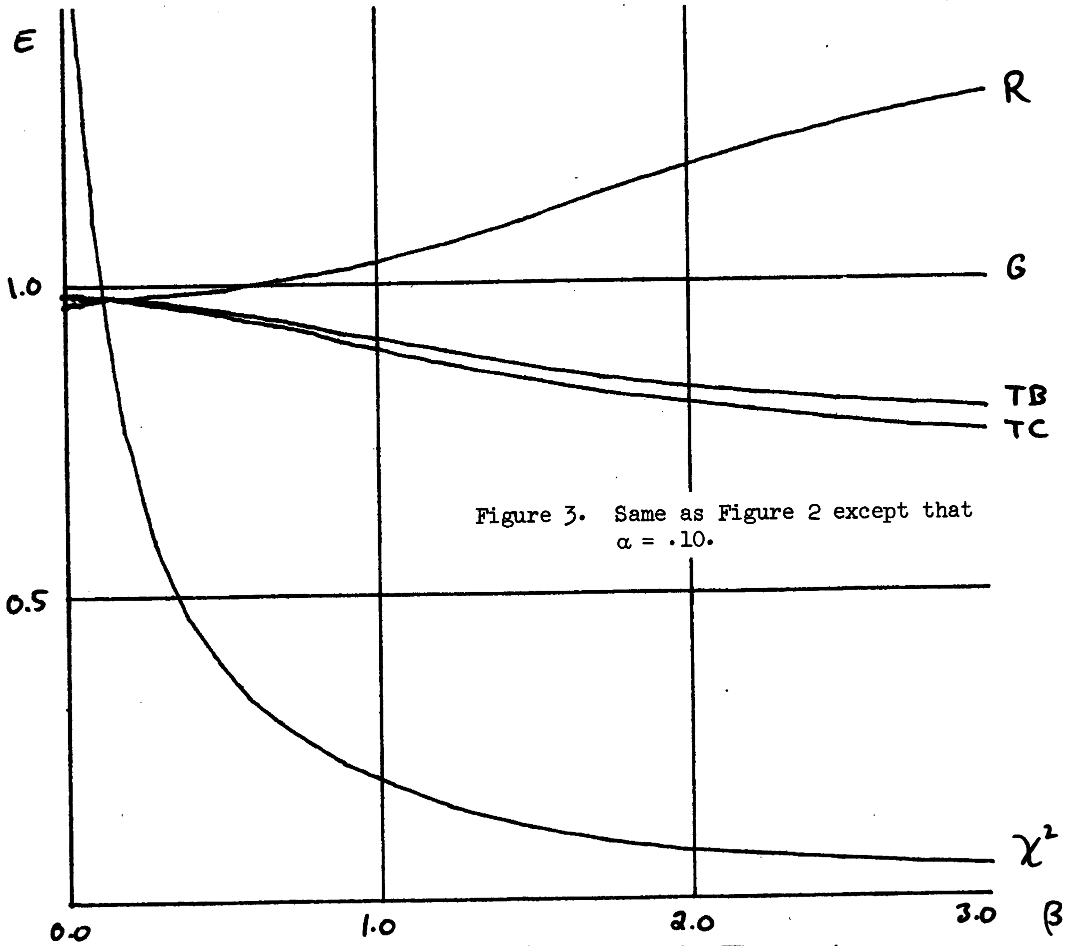

Figure 3. Same as Figure 2 except that $\alpha = .10$.

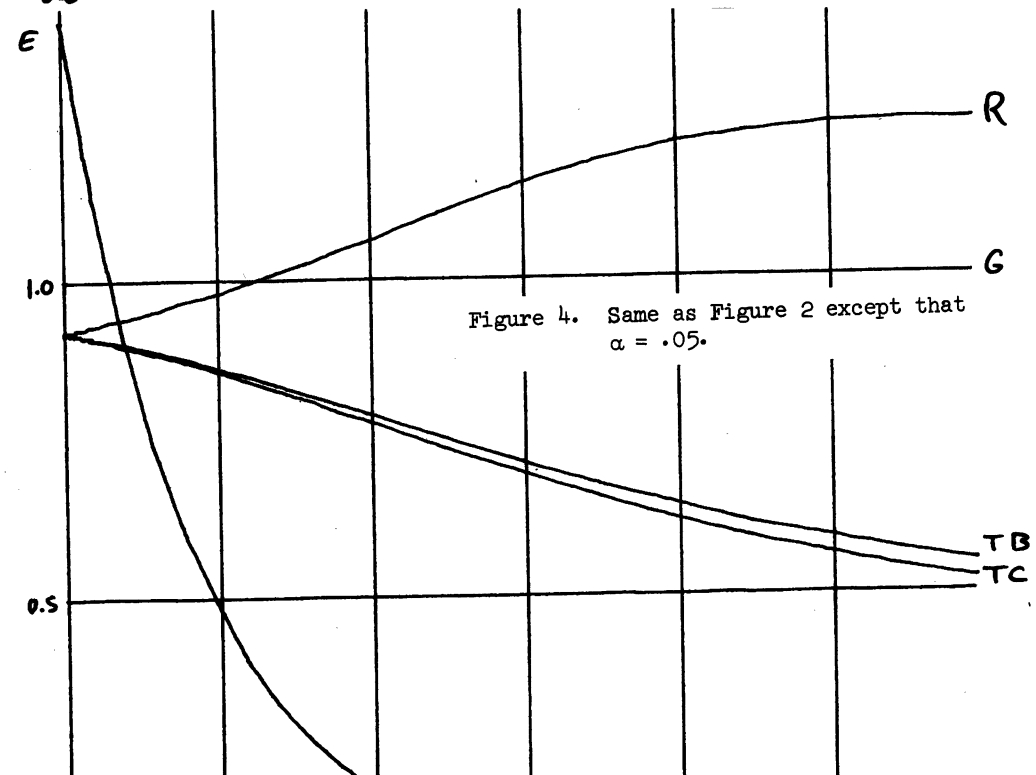

Figure 4. Same as Figure 2 except that $\alpha = .05$.

tables below. As a precision for the chi-square test to be put into (5.5) we will use:

$$(6.4)\quad P_{\chi^2} = 7.85\ \lambda/\lambda(DF, .05, .80) .$$

7. Numerical Results on Efficiencies

In presenting results on relative efficiencies the coefficient G has been taken as a base of comparison and its efficiency set at 1.000. The relative efficiencies of the other statistics become $E_{TC} = P_C/P_G$, $E_{TB} = P_B/P_G$, $E_R = P_R/P_G$ and $E_{\chi^2} = P_{\chi^2}/P_G$. For the values of the π_{ij}, probabilities under the joint normal distribution were used in the first results, and A and B were both taken as six. The π_{ij} were probabilities in a joint distribution with correlation coefficient equal to .80 and the marginal probabilities were all set equal to one-sixth.[3] The matrix of the π_{ij} thus appears as:

$$(\pi_{ij}) = \begin{pmatrix} .106 & .031 & .011 & .010 & .004 & .004 \\ .031 & .074 & .034 & .018 & .005 & .004 \\ .011 & .034 & .063 & .031 & .018 & .010 \\ .010 & .018 & .031 & .063 & .034 & .011 \\ .004 & .005 & .018 & .034 & .074 & .031 \\ .004 & .004 & .010 & .011 & .031 & .106 \end{pmatrix} .$$

If there is no error, that is $\rho_{ab} = \pi_{ij}$, then the following are the efficiencies:

$$E_G = 1.000,\ E_{TC} = .825,\ E_{TB} = .819$$
$$E_R = 1.006,\ E_{\chi^2} = .271 .$$

One concludes for this case of exact measurement of an interval-type association that R, G, TB and TC have nearly the same sensitivity to such an alternative hypothesis, while chi-square is much less sensitive.

As measurement or drift error is introduced by way of the parameter α, we find, in general, that R, TB and TC drop in efficiency relative to G when we maintain the underlying joint normal distribution as a source of π_{ij} values. Figure 1 shows this for $\alpha = .15$. Notice how similar are TC and TB, as might be expected. Also notice how R recovers its efficiency as intervalness of the error pattern increases, _i.e._, as β increases.

When the error-free distribution is taken to be a diagonal of one-sixth probabilities then the results shown in Figure 2 appear. In this case the coefficient R becomes the most efficient as intervalness increases. TB is a bit more efficient than TC but both are somewhat below G. The pattern of results in Figure 3 can be accentuated by reducing the level of error from $\alpha = .15$ to $\alpha = .10$ and $\alpha = .05$. These results are shown in Figures 4 and 5.

8. Conclusions

A major conclusion is that, 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. This seems to suggest that in the absence of any clear-cut superior measure of association the data analyst should use whichever measure he favors. However, he should direct some attention to the measurement error process. There should be retesting and re-coding done so as to learn between which categories is there misclassification. Then this source of uncertainty can be adjusted for in estimating the pattern of association among the true-category variates. Also, when working with qualitative data, one should consider formulating a realistic probability model for the cell frequencies and then estimating the parameters of that model rather than estimating some conventionally employed measure of association.

If, however, a measure of association must be recommended then my judgement would be the following. In so far as it seems reasonable that a numerical variate will be defined some day and the present categories will become ranges of this variate, then use scores and compute R. If not, then use G when two variates are involved, particularily if it seems that the measurement errors follow a relatively "flat" pattern with a small β. Kendall and Stuart [13, p. 566] report results suggesting the superiority of G over TC and conclude that: "If this is shown to be true in general, this fact ... would make [G] likely to become the standard measure of association for the ordered case." Our calculations show it to be true fairly generally. If more than two categorical variates with measurement error pattern having small β are being analyzed then one might use TB. This is because TB is a bona fide correlation coefficient, albeit of pair data. Consequently, its correlation matrix will be positive semi-definite and so lends itself to multivariate calculations.

FOOTNOTES

1. Although G, TB, and R are ratios of random variables, TC is not; and it is possible to obtain an expression for the exact variance of TC as:

$$(5.4a)\quad V(TC) = \left(\frac{A}{A-1}\right)^2 \frac{4}{n-1} \times [\Pi_{dd} - 2\Pi_{sd} + \Pi_{dd} - (\Pi_s - \Pi_d)^2] + \frac{2}{n(n-1)} [\Pi_s + \Pi_d - 4(\Pi_{dd} - 2\Pi_{sd} + \Pi_{ss}) + 3(\Pi_s - \Pi_d)^2]$$

When n is of moderate size, one would multiply TC by (n-1)/n to obtain the quantity t_c in

[13, p. 563], and then the variance expression (5.4a) would be multiplied by $[(n-1)/n]^2$.

2. It may be of interest to note how the proportional importance of the extra part of formula (5.4) to the part in 1/n decreases. For a case with $\tau_c = .247$ in a 6 by 6 table the exact variance was found to be $.42878/n-1 + .88073/n(n-1)$. This suggests that for moderate sample sizes our approximation would not be too good, but for large sample sizes the difference is negligible. With n = 20, the increase of exact over approximate variance is 9.3% and for n = 100 it drops to 2.0%.

3. These calculations were made by Anne Dalhouse using the charts provided in the National Bureau of Standard's Handbook [1, pp. 936].

REFERENCES

[1] Abramowitz, M. and Stegun, I. A. (editors). Handbook of Mathematical Functions, National Bureau of Standards, Applied Mathematics Series, No. 55, Washington, D. C.: U. S. Government Printing Office, 1964.

[2] Assakul, K. and Proctor, C. H. "Testing Independence in Two-way Contingency Tables with Data Subject to Misclassification," Psychometrika, 32 (March, 1967), 67-76.

[3] Bahadur, R. R. "Stochastic Comparison of Tests," The Annals of Mathematical Statistics, 31 (June, 1960), 276-295.

[4] Barnard, G. A. "Significance Tests for 2 × 2 Tables," Biometrika, 34 (January, 1947), 123-

[5] Campbell, N. R. Foundations of Science: The Philosophy of Theory and Experiment, New York: Dover Publications, Inc., 1957.

[6] Costner, H. L. "Criteria for Measures of Association," American Sociological Review, 30 (June, 1965), 341-353.

[7] Farlie, D. J. G. "The Asymptotic Efficiency of Daniels' Generalized Correlation Coefficients," Journal of the Royal Statistical Society, Series B 23, No. 1 (1961), 128-142.

[8] Fisher, R. A. Statistical Methods for Research Workers, 13th edition-revised, New York: Hafner Publishing Company, Inc., 1958.

[9] Fix, E. "Tables of Noncentral X^2," University of California Publication in Statistics, 1 (1949), 15-19.

[10] Goodman, L. A. and Kruskal, W. H. "Measures of Association for Cross Classification, II: Further Discussion and References," Journal of the American Statistical Association, 54 (March, 1959), 123-163.

[11] Goodman, L. A. and Kruskal, W. H. "Measures of Association for Cross Classification III: Approximate Sampling Theory," Journal of the American Statistical Association, 58 (June, 1963), 310-364.

[12] Hayashi, C. "Response Errors and Biased Information," Annals of the Institute of Statistical Mathematics," 20 (No. 2, 1968), 24-28.

[13] Kendall, M. G. and Stuart, Alan. The Advanced Theory of Statistics: Vol. 2 Inference and Relationship, New York: Hafner Publishing Company, 1961.

[14] Konijn, H. S. "On the Power of Certain Tests for Independence in Bivariate Populations," The Annals of Mathematical Statistics, 27 (June, 1956), 300-323.

[15] Kruskal, W. H. "Ordinal Measures of Association," Journal of the American Statistical Association, 53 (December, 1958), 814-861.

[16] Morris, R. N. "Multiple Correlation and Ordinally Sealed Data," Social Forces, (March, 1970), 299-311.

[17] Plackett, R. L. "A Class of Bivariate Distributions," Journal of the American Statistical Association, 60 (June, 1965), 516-522.

[18] Mote, V. L. and Anderson, R. L. "The Effect of Misclassification on the Properties of X^2 - Tests," Biometrika, 52 (June, 1965), 95-109.

[19] Mosteller, F. "Association and Estimation in Contingency Tables," Journal of the American Statistical Association, 63 (March, 1968), 1-28.

[20] Rao, C. R. Linear Statistical Inference and Its Applications, New York: John Wiley and Sons, Inc., 1965.

[21] Suppes, P. and Zinnes, J. L. "Basic Measurement Theory," Chapter 1 in Luce, R. D., Bush, R. R. and Galanter, E. (editors), Handbook of Mathematical Psychology, Vol. 1, New York: John Wiley and Sons, Inc., 1963.

[22] Williams, E. J. "Use of Scores for the Analysis of Association in Contingency Tables," Biometrika, 39 (December, 1952), 274-289.

[23] Yates, F. "The Analysis of Contingency Tables with Groupings Based on Qualitative Characters," Biometrika, 35 (May, 1948), 176-181.