

Stuart's tau measure of effect size for ordinal variables: Some methodological considerations

KENNETH J. BERRY

Colorado State University, Fort Collins, Colorado

JANIS E. JOHNSTON

AAAS Science and Technology Policy Fellow, Washington, D.C.

AND

SAMMY ZAHARAN AND PAUL W. MIELKE, JR.

Colorado State University, Fort Collins, Colorado

The reporting of measures of effect size has become increasingly important in psychology. A Monte Carlo resampling permutation procedure is introduced to find near-optimum maximum values for Stuart's τ_c measure for two-way ordinal contingency tables, also termed Kendall's τ_c since Kendall introduced τ_a and τ_b . Comparisons between resampling and exact procedures demonstrate the accuracy and utility of resampling measures of effect size for two-way ordinal contingency tables. The resampling procedure is shown to be more precise than the traditional method of standardizing τ_c .

Hays (1963) challenged psychologists to report measures of effect size in addition to the usual tests of significance. Later in the decade, Vaughan and Corballis (1969) lamented the inattention paid to the problem described by Hays. The fourth edition of the *Publication Manual of the American Psychological Association* (APA; 1994) suggested that tests of statistical significance be accompanied by measures of effect size. That language was strengthened with the publication of the fifth edition of the *Publication Manual of the American Psychological Association* (2001) and a call to make measures of effect size an integral part of publishing in APA journals. Consequently, a number of academic journals, both APA and others, began requiring measures of effect size as a condition of publication (Capraro & Capraro, 2002; Rosnow, Rosenthal, & Rubin, 2000; Thompson, 2002; Vacha-Haase, Nilsson, Reetz, Lance, & Thompson, 2000). Measures of effect size provide essential information about the substantive import of experimental results. Thus, it is important to explore new measures of effect size while continuing to reassess and refine existing measures.

It is often important in psychological research to analyze the relationship between two variables measured on ordinal scales, such as Likert scales ("Strongly Agree," . . . , "Strongly Disagree"), letter grades (A, B, . . . , F), self-ratings (pain, depression, etc.), meteorological ratings (tornadoes, tropical cyclones, etc.), socioeconomic status (upper, middle, lower), power/influence (high, medium, low), prestige (journals, celebrities, colleges, etc.), rankings (academic, military, organizational, etc.), orders

(birth order, order of finish in a race, biological taxonomy, etc.), and psychological scales (alienation, anxiety, authoritarianism, anomie, etc.). Indeed, it has long been recognized that "most of the scales used . . . by psychologists are ordinal scales" (Stevens, 1946, p. 679).

Kendall (1938, 1945) developed the τ_a and τ_b measures of rank correlation that are now widely used as measures of effect size when two ordinal variables are cross-tabulated into an $r \times c$ contingency table, where r and c denote the number of rows and columns, respectively. Stuart (1953) introduced a third measure of rank correlation, τ_c , which in a classic case of Stigler's (1980) law of eponymy is often erroneously referred to as Kendall's τ_c . Stuart's τ_c measure of ordinal association was designed to replace Kendall's τ_a and τ_b and was specifically designed for ordinal contingency tables where $r \neq c$ and some values are tied. Under many circumstances, τ_c reduces to τ_b when $r = c$ and always reduces to τ_a when $r = c$ and no values are tied. When $r \neq c$, τ_c may not properly norm between -1 and $+1$. This article introduces a Monte Carlo resampling method to properly norm Stuart's τ_c .

Measures of Effect Size

In recent years, there has been increased emphasis on reporting measures of effect size in addition to tests of significance, recognizing that determination of a significant treatment effect does not necessarily translate into a substantial treatment effect (Hunter, 1997; Killeen, 2005; Loftus, 1996). The increased attention to effect size highlighted two problems: First, measures of effect size do

K. J. Berry, berry@mail.colostate.edu

not exist for all tests of significance (Johnston, Berry, & Mielke, 2004); second, measures of effect size that are in routine use may over- or underestimate the real effect size.

Stuart's τ_c is based on an unstandardized measure of effect size, commonly termed S . Consider n bivariate observations. Let (x_i, y_i) and (x_j, y_j) represent a pair of observations, $i, j = 1, \dots, n$. If $(x_j - x_i)$ and $(y_j - y_i)$ have the same sign, the pair is concordant; if the signs are opposite, the pair is discordant. A simple way to measure strength of relationship is to compute the number of concordant pairs minus the number of discordant pairs, S , ignoring all tied x and y values. A predominance of concordant pairs results in a large positive value of S and indicates a strong positive relationship between x and y . A predominance of discordant pairs results in a large negative value of S and indicates a strong negative relationship between x and y . As a measure of effect size—that is, strength of relationship— S has a distinct disadvantage. Its range depends on the sample size n and needs to be standardized so that it norms between -1 and $+1$.

Several possibilities exist to bound S between -1 and $+1$, representing complete dissociation and complete association of x and y , respectively. Stuart (1953) proposed the maximum value of S be defined as $n^2(m - 1)/(2m)$, where $m = \min(r, c)$ and provided $n \bmod m = 0$. Given the maximum value of S proposed by Stuart, τ_c is a generalized measure of effect size for two cross-classified ordinal variables, where $\tau_c = (2mS)/[n^2(m - 1)]$.

In order for S to equal $n^2(m - 1)/(2m)$ and, thus, for τ_c to attain the maximum value when $-1 \leq \tau_c \leq 1$, two conditions must be met. For simplicity, consider an $r \times c$ contingency table with $r \leq c$, then $r = \min(r, c)$ and $c = \max(r, c)$. First, each row marginal must equal n/m , implying $n \bmod m = 0$ (Stuart, 1953). Second, some sum of 1 to c column marginals must equal n/m , summing sequentially either from left to right, beginning with the first column, or from right to left, beginning with column c .

Thus, the problem with τ_c lies in the definition for the maximum value for S provided by Stuart (1953). The denominator, $n^2(m - 1)/(2m)$, provides only an upper bound for S computed on an idealized $r \times c$ contingency table. In most cases, however, $n^2(m - 1)/(2m)$ yields a higher value than the maximum possible value of S calculated on the observed contingency table. Consequently, the effect size is often underestimated. For an $r \times c$ table with $r \leq c$, $n^2(m - 1)/(2m)$ is always equal to $n^2(r - 1)/(2r)$, since it is based solely on n and $\min(r, c)$. The solution is to find a sharper bound for the maximum value of τ_c than Stuart's $n^2(m - 1)/(2m)$ provides.

Resampling

An alternative to Stuart's $n^2(m - 1)/(2m)$ is provided by a Monte Carlo resampling algorithm. Given two integral marginal vectors, the set of all $r \times c$ contingency tables with row marginals $\{n_{1+}, \dots, n_{r+}\}$ and column marginals $\{n_{+1}, \dots, n_{+c}\}$ is a Fréchet class of matrices of nonnegative integer elements given fixed marginals and denoted by $F(a, b)$ (Fréchet, 1951). Enumerating all M members of

$F(a, b)$ to find an exact solution is computationally prohibitive and impractical, since the reference set, M , is usually very large even for modest values of r and c . An alternative to an exact solution is a Monte Carlo resampling algorithm that enumerates a random sample of size L from all the M members of $F(a, b)$. If T denotes the statistic of interest, a resampling algorithm is constructed as follows.¹

1. Let L denote a random sample with replacement of a large number of the possible arrangements of the n_{ij} cell frequencies for $i = 1, \dots, r$ and $j = 1, \dots, c$ with fixed marginals $\{n_{1+}, \dots, n_{r+}\}$ and $\{n_{+1}, \dots, n_{+c}\}$.

2. Set counter k and the maximum value of statistic T (T_{\max}) to zero.

3. If each $\{n_{1+}, \dots, n_{m+}\}$ or $\{n_{+1}, \dots, n_{+m}\}$ marginal corresponding to $m = \min(r, c)$ equals n/m , go to 4; otherwise, go to 5.

4. Set $w = \max(r, c)$. If $w = c$ and any sequence of marginals beginning with column 1 or column w sums to n/m , or if $w = r$ and any sequence of marginals beginning with row 1 or row w sums to n/m , then set $T_{\max} = n^2(m - 1)/(2m)$ and go to 9; otherwise, go to 5.

5. Generate a random arrangement of the n_{ij} cell frequencies for $i = 1, \dots, r$ and $j = 1, \dots, c$ satisfying the fixed marginals $\{n_{1+}, \dots, n_{r+}\}$ and $\{n_{+1}, \dots, n_{+c}\}$. An algorithm by Patefield (1981) ensures random arrangements of the n_{ij} cell frequencies, given the fixed marginals.

6. Compute statistic T on the random arrangement of the n_{ij} values and set counter $k = k + 1$.

7. If $T > T_{\max}$, T_{\max} is replaced by T .

8. If $k = L$, the maximum resampled value of T is T_{\max} , go to 9; otherwise, go to 5.

9. Exit

Stuart's τ_c examples. Determination of the exact maximum value of S is impractical for many $r \times c$ contingency tables. However, generation of all M arrangements of $F(a, b)$ is possible when r and c are small (cf. Mielke & Berry, 1992). Table 1 contains maximum S values utilizing exact, resampling, and Stuart's procedures for $2 \times 2, 2 \times 3, 2 \times 4, 2 \times 5, 2 \times 6, 3 \times 3, 3 \times 4$, and 4×4 contingency tables. Because the true maximum value of S is dependent on the marginal frequency distributions, a variety of uniform and skewed marginal frequency distributions are utilized in Table 1. Uniform marginal frequency distributions imply that the probability of an observation falling into row i is $1/r$ and the probability of an observation falling into column j is $1/c$. On the other hand, the skewed marginal frequency distributions in Table 1 imply that the probability of an observation falling into row i is $(2i)/[r(r + 1)]$ and the probability of an observation falling into column j is $(2j)/[c(c + 1)]$. The various values of $n = 18, 36, 60, 90$, and 126 for the 29 contingency tables in Table 1 were obtained from $n = 3\max[r(r + 1), c(c + 1)]$, thus ensuring integral values for the $r = 2, 3, 4$ and $c = 2, 3, 4, 5, 6$ marginals. The "Exact" column in Table 1 lists maximum values of S based on all M members of $F(a, b)$; the "Resampling" column lists maximum values of S based on $L = 1,000,000$ resampled arrangements of cell frequencies, given fixed marginals; and the "Stuart" column lists maximum values of S based on $n^2(m - 1)/(2m)$.

Table 1
Maximum *S* Values Based on Exact, Resampling, and Stuart's Procedures for
2×2, 2×3, 2×4, 2×5, 2×6, 3×3, 3×4, and 4×4 Contingency Tables With
Uniform and Skewed Row and Column Marginal Frequency Totals

| Table Number | Table Size | Marginal | | Procedure | | | |
|--------------|------------|------------------|--------------------------|--------------------------|------------|--------|-------|
| | | Row | Column | Exact | Resampling | Stuart | |
| 1 | 2×2 | {9, 9} | {9, 9} | 81 | 81 | 81 | |
| 2 | (n = 18) | {9, 9} | {6, 12} | 54 | 54 | 81 | |
| 3 | | {6, 12} | {6, 12} | 72 | 72 | 81 | |
| 4 | | {18, 18} | {12, 12, 12} | 288 | 288 | 324 | |
| 5 | (n = 36) | {18, 18} | {6, 12, 18} | 324 | 324 | 324 | |
| 6 | | {12, 24} | {12, 12, 12} | 288 | 288 | 324 | |
| 7 | | {12, 24} | {6, 12, 18} | 252 | 252 | 324 | |
| 8 | 2×4 | {30, 30} | {15, 15, 15, 15} | 900 | 900 | 900 | |
| 9 | | (n = 60) | {30, 30} | {6, 12, 18, 24} | 828 | 828 | 900 |
| 10 | | | {20, 40} | {15, 15, 15, 15} | 750 | 750 | 900 |
| 11 | {20, 40} | | {6, 12, 18, 24} | 768 | 708 | 900 | |
| 12 | 2×5 | {45, 45} | {18, 18, 18, 18, 18} | 1,944 | 1,944 | 2,025 | |
| 13 | | (n = 90) | {45, 45} | {6, 12, 18, 24, 30} | 1,890 | 1,764 | 2,025 |
| 14 | | | {30, 60} | {18, 18, 18, 18, 18} | 1,728 | 1,728 | 2,025 |
| 15 | {30, 60} | | {6, 12, 18, 24, 30} | 1,728 | 1,728 | 2,025 | |
| 16 | 2×6 | {63, 63} | {21, 21, 21, 21, 21, 21} | 3,969 | 3,969 | 3,969 | |
| 17 | | (n = 126) | {63, 63} | {6, 12, 18, 24, 30, 36} | 3,888 | 3,042 | 3,969 |
| 18 | | | {42, 84} | {21, 21, 21, 21, 21, 21} | 3,528 | 3,528 | 3,969 |
| 19 | {42, 84} | | {6, 12, 18, 24, 30, 36} | 3,420 | 3,420 | 3,969 | |
| 20 | 3×3 | {12, 12, 12} | {12, 12, 12} | 432 | 432 | 432 | |
| 21 | | (n = 36) | {12, 12, 12} | {6, 12, 18} | 324 | 324 | 432 |
| 22 | | | {6, 12, 18} | {6, 12, 18} | 396 | 396 | 432 |
| 23 | 3×4 | | {20, 20, 20} | {15, 15, 15, 15} | 1,100 | 1,100 | 1,200 |
| 24 | (n = 60) | {20, 20, 20} | {6, 12, 18, 24} | 1,088 | 876 | 1,200 | |
| 25 | | {10, 20, 30} | {15, 15, 15, 15} | 1,050 | 1,050 | 1,200 | |
| 26 | | {10, 20, 30} | {6, 12, 18, 24} | 996 | 996 | 1,200 | |
| 27 | 4×4 | {15, 15, 15, 15} | {15, 15, 15, 15} | 1,350 | 1,350 | 1,350 | |
| 28 | | (n = 60) | {15, 15, 15, 15} | {6, 12, 18, 24} | 1,116 | 1,035 | 1,350 |
| 29 | | | {6, 12, 18, 24} | {6, 12, 18, 24} | 1,260 | 1,260 | 1,350 |

As is evident in Table 1, the maximum *S* values based on resampling are always less than those obtained by the maximum proposed by Stuart (1953) except in six cases: the 2×2 table with marginals of {9, 9} and {9, 9}; the 2×3 table with marginals of {18, 18} and {6, 12, 18}; the 2×4 table with marginals of {30, 30} and {15, 15, 15, 15}; the 2×6 table with marginals of {63, 63} and {21, 21, 21, 21, 21, 21}; the 3×3 table with marginals of {12, 12, 12} and {12, 12, 12}; and the 4×4 table with marginals of {15, 15, 15, 15} and {15, 15, 15, 15}. All six cases satisfy the two conditions for $n^2(m - 1)/(2m)$ —that is, row marginals equal to n/m and some sequential sum of column marginals equal to n/m .

Proper comparisons are between the resampled maximum values of *S* and Stuart's maximum values of *S*. The purpose of these analyses is to obtain sharper bounds on maximum *S* through resampling; the exact values of *S* are listed in Table 1 only to demonstrate optimal results, and for the example analyses in Table 1 constitute a gold standard for purposes of comparison.

Stuart's procedure matches the exact maximum value of *S* for only 6 of the 29 marginal conditions specified in Table 1. Thus, Stuart's procedure often overestimates the maximum value of *S* and, consequently, underestimates effect size. As can be seen in Table 1, there are five tables where the resampling and exact procedures differ on the maximum value of *S*: 11, 13, 17, 24, and 28. Note that

all five tables have similar skewed column marginal frequency distributions of {6, 12, 18, 24}, {6, 12, 18, 24, 30}, or {6, 12, 18, 24, 30, 36}. Consequently, the number of possible cell frequency configurations yielding the maximum value of *S* is severely circumscribed, other factors being equal. Table 2 lists the table numbers from Table 1 for those tables where the resampling and exact maximum values of *S* did not agree, the exact number of possible cell frequency configurations given the fixed marginals, the exact maximum value of *S*, the point probability value of each table yielding the maximum value of *S*, and a listing of the cell frequency configurations yielding the maximum value of *S*.

To illustrate, consider table number 11, in Table 2, for which there are $M = 1,088$ possible cell frequency configurations given the fixed marginals of {20, 40} and {6, 12, 18, 24}, the maximum value of *S* is 768, the point probability value is only 0.3650×10^{-13} , and only one of the $M = 1,088$ possible cell configurations yields a maximum *S* value of 768. Note that for contingency table 13 in Table 2, there are two cell frequency configurations yielding a maximum value of 1,890, because the row marginal frequencies are both equal to $n/m = 45$. For the five tables listed in Table 2, it is not surprising that the exact and resampling values of *S* differ, given $L = 1,000,000$, the large number of possible cell frequency configurations, the skewed marginal frequency distributions, the limited number(s) of cell

Table 2
Tables 11, 13, 17, 24, and 28 From Table 1 With Total Number of Possible Cell Frequency Configurations, Maximum *S* Values, and Probability Values and Cell Frequency Configurations Corresponding to the Maximum *S* Value

| Table Number | Number of Configurations | Maximum <i>S</i> Value | Point Probability | Cell Frequencies |
|--------------|--------------------------|------------------------|--|--|
| 11 | 1,088 | 768 | 0.3650×10^{-13} | 6 12 2 0 0 0 16 24 |
| 13 | 37,775 | 1,890 | 0.1259×10^{-19} 0.1259×10^{-19} | 6 12 18 9 0 0 0 0 15 30 0 0 0 15 30 6 12 18 9 0 |
| 17 | 1,102,967 | 3,888 | 0.6727×10^{-33} | 6 12 18 24 3 0 0 0 0 0 27 36 |
| 24 | 358,267 | 1,088 | 0.2814×10^{-20} | 6 12 2 0 0 0 16 4 0 0 0 20 |
| 28 | 28,904,292 | 1,116 | 0.1877×10^{-20} | 6 9 0 0 0 3 12 0 0 0 6 9 0 0 0 15 |

frequency configurations yielding the maximum value of *S*, and the very small point probability values of the cell frequency configuration(s) yielding the maximum value of *S*.

Although it is constructive to specify under what conditions the resampling and exact procedures sometimes lead to different results, it is important to bear two things in mind: First, exact solutions are impractical for most applications and are included here only to establish the accuracy of the resampling procedure with $L = 1,000,000$; second, the proper comparison to be made in Table 1 is between the resampling and Stuart procedures, where the resampling procedure is clearly superior in finding the maximum value of *S*.

Although Table 1 documents the possible limitations of Stuart’s proposed maximum value of *S*, the question remains as to the effect of the maximum value on the standardized measure of effect size, τ_c . It is obvious that different maximum values of *S* obtained by the resampling and Stuart’s procedures have little effect on the value of τ_c when the observed value of *S* is zero or close to zero. Moreover, researchers care little about small effect sizes. Table 3 lists four 2×2, four 2×3, and four 2×4 contingency tables, with the column marginals, observed cell frequencies, maximum values of *S* based on resampling, maximum values of *S* based on Stuart’s $n^2(m - 1)/(2m)$, observed values of τ_c based on the resampled maximum

Table 3
Example 2×2, 2×3, and 2×4 Contingency Tables With $n = 60$, Row Marginals of {30, 30}, and Observed Cell Frequencies Chosen to Maximize the Observed Value of *S*

| Table Size | Column Marginals | Observed Cell Frequencies | Maximum <i>S</i> Value | | Observed τ_c Value | |
|------------|------------------|---------------------------|------------------------|--------|-------------------------|--------|
| | | | Resampling | Stuart | Resampling | Stuart |
| 2×2 | {30, 30} | 30 0 0 30 | 900 | 900 | +1.000 | +1.000 |
| | | {20, 40} | 20 10 0 30 | 600 | 900 | +1.000 |
| | {10, 50} | 10 20 0 30 | 300 | 900 | +1.000 | +0.333 |
| | | {5, 55} | 5 25 0 30 | 150 | 900 | +1.000 |
| 2×3 | {10, 20, 30} | 10 20 0 0 0 30 | 900 | 900 | +1.000 | +1.000 |
| | | {20, 20, 20} | 20 10 0 0 10 20 | 800 | 900 | +1.000 |
| | {5, 20, 35} | 5 20 5 0 0 30 | 750 | 900 | +1.000 | +0.833 |
| | | {5, 15, 40} | 5 15 10 0 0 30 | 600 | 900 | +1.000 |
| 2×4 | {15, 15, 15, 15} | 15 15 0 0 0 0 15 15 | 900 | 900 | +1.000 | +1.000 |
| | | {10, 15, 15, 20} | 10 15 5 0 0 0 10 20 | 850 | 900 | +1.000 |
| | {5, 15, 15, 25} | 5 15 10 0 0 0 5 25 | 850 | 900 | +1.000 | +0.944 |
| | | {5, 10, 10, 35} | 5 10 10 5 0 0 0 30 | 750 | 900 | +1.000 |

value of S , and observed values of τ_c based on Stuart's maximum value of S for each of the 12 tables.

In order to isolate the effect of skewed marginals on the value of τ_c , each of the 12 tables in Table 3 has $n = 60$, $r = 2$ rows, identical uniform row marginals of $n/m = 60/2 = \{30, 30\}$, and observed cell frequencies designed to ensure a maximum value of S , thus controlling for n , r , n_{i+} for $i = 1, \dots, 2$, and S . The fourth and fifth columns in Table 3 list the maximum values of S obtained by the resampling and Stuart procedures. It should be noted that the resampling and exact maximum values of S are the same in these examples. Since Stuart's procedure is based solely on $n = 60$ and $m = 2$, the maximum value is identical for all 12 tables. The last two columns of Table 3 list the observed values of τ_c on the basis of the resampling and Stuart procedures. Comparison of the last two columns reveals that, whereas differences between the two procedures are at times nonexistent or insignificant, at other times the differences can be quite large, due to skewed column marginal distributions that neither equal n/m nor sum sequentially to one of the row marginals, the most extreme example being the 2×2 contingency table with highly skewed column marginals of $\{5, 55\}$, where τ_c is $+1.000$ under the resampling procedure and only $+0.167$ under the Stuart procedure.

Discussion

The resampling method for calculating sharper bounds for τ_c is an example of a relatively new technique, Monte Carlo resampling, applied to an existing statistic that enables improvement in measurement accuracy. The Monte Carlo resampling permutation procedure provides sharper bounds for the maximum value of S , permitting better estimation of effect sizes than can be accomplished with Stuart's maximum value of S based on $n^2(m - 1)/(2m)$. Table 1 demonstrates the effectiveness of resampling in providing sharper bounds for maximum S than Stuart's $n^2(m - 1)/(2m)$ over a variety of table sizes, sample sizes, and marginal distributions. Stuart's procedure systematically deflates effect sizes by overestimating the maximum value of S in 23 of the 29 marginal conditions specified. Table 2 provides some rationale for those instances when the exact and resampling maximum values of S do not agree. Finding the maximum value of S with resampling is not a simple matter of finding one out of M possible cell frequency configurations, since the probabilities of different configurations differ. In general, the success of the resampling procedure depends on the size of M , the skewness of the marginal frequency distributions, the number of cell frequency configurations yielding the maximum value of S , and the point probability of the cell frequency configuration(s) yielding the maximum value of S . Table 3 explores the impact of the wrong maximum value of S on the value of τ_c , while controlling for n , r , n_{i+} for $i = 1, \dots, r$, and S . Inspection of Table 3 reveals that skewed marginals often lead to inflated values of S and generate τ_c values that are too small, sometimes by a substantial amount.

AUTHOR NOTE

Correspondence concerning this article should be addressed to K. J. Berry, Department of Sociology, Colorado State University, Fort Collins, CO 80523-1784 (e-mail: berry@mail.colostate.edu).

REFERENCES

- AMERICAN PSYCHOLOGICAL ASSOCIATION (1994). *Publication manual of the American Psychological Association* (4th ed.). Washington, DC: Author.
- AMERICAN PSYCHOLOGICAL ASSOCIATION (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.
- CAPRARO, R. M., & CAPRARO, M. M. (2002). Treatments of effect sizes and statistical significance in textbooks. *Educational & Psychological Measurement*, *62*, 771-782.
- FRÉCHET, M. (1951). Sur les tableaux de corrélation dont les marges sont donnés. *Annales de l'Université de Lyon*, *4*, 53-77.
- HAYS, W. L. (1963). *Statistics for psychologists*. New York: Holt, Rinehart & Winston.
- HUNTER, J. E. (1997). Needed: A ban on the significance test. *Psychological Science*, *8*, 3-7.
- JOHNSTON, J. E., BERRY, K. J., & MIELKE, P. W., JR. (2004). A measure of effect size for experimental designs with heterogeneous variances. *Perceptual & Motor Skills*, *98*, 3-18.
- KENDALL, M. G. (1938). A new measure of rank correlation. *Biometrika*, *30*, 81-93.
- KENDALL, M. G. (1945). The treatment of ties in ranking problems. *Biometrika*, *33*, 239-251.
- KILLEEN, P. R. (2005). An alternative to null-hypothesis significance tests. *Psychological Science*, *16*, 345-353.
- LOFTUS, G. R. (1996). Psychology will be a much better science when we change the way we analyze data. *Current Directions in Psychological Science*, *5*, 161-171.
- MIELKE, P. W., & BERRY, K. J. (1992). Fisher's exact probability test for cross-classification tables. *Educational & Psychological Measurement*, *52*, 97-101.
- PATEFIELD, W. M. (1981). Algorithm AS159: An efficient method of generating random $R \times C$ tables with given row and column totals. *Applied Statistics*, *30*, 91-97.
- ROSNOW, R. L., ROSENTHAL, R., & RUBIN, D. B. (2000). Contrasts and correlation in effect-size estimation. *Psychological Science*, *11*, 446-453.
- STEVENS, S. S. (1946). On the theory of scales of measurement. *Science*, *103*, 677-680.
- STIGLER, S. (1980). Stigler's law of eponymy. In T. F. Gieryn (Ed.), *Science and social structure: A festschrift for Robert K. Merton* (pp. 147-157). New York: New York Academy of Sciences.
- STUART, A. (1953). The estimation and comparison of strengths of association in contingency tables. *Biometrika*, *40*, 105-110.
- THOMPSON, B. (2002). "Statistical," "practical," and "clinical": How many kinds of significance do counselors need to consider? *Journal of Counseling & Development*, *80*, 64-71.
- VACHA-HAASE, T., NILSSON, J. E., REETZ, D. R., LANCE, T. S., & THOMPSON, B. (2000). Reporting practices and APA editorial policies regarding statistical significance and effect size. *Theory & Psychology*, *10*, 413-425.
- VAUGHAN, G. M., & CORBALLIS, M. C. (1969). Beyond tests of significance: Estimating strength of effects in selected ANOVA designs. *Psychological Bulletin*, *72*, 204-213.

NOTE

1. The FORTRAN program STC is available from the first author at the correspondence address above.

(Manuscript received February 16, 2009;
revision accepted for publication August 9, 2009.)