

Note on Rank-Biserial Correlation when There Are Ties

José Moral de la Rubia

School of Psychology, Universidad Autónoma de Nuevo León, Monterrey, México

Email: jose.moral@uanl.edu.mx, jose_moral@hotmail.com

How to cite this paper: de la Rubia, J.M. (2022) Note on Rank-Biserial Correlation when There Are Ties. *Open Journal of Statistics*, 12, 597-622.
<https://doi.org/10.4236/ojs.2022.125036>

Received: August 26, 2022

Accepted: October 10, 2022

Published: October 13, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The objective of this article is to demonstrate with examples that the *two-sided tie* correction does not work well. This correction was developed by Cureton so that Kendall's tau-type and Spearman's rho-type formulas for rank-biserial correlation yield the same result when ties are present. However, a correction based on the *bracket ties* achieves the desired goal, which is demonstrated algebraically and checked with three examples. On the one hand, the 10-element random sample given by Cureton, in which the *two-sided tie* correction performs well, is taken up. On the other hand, two other examples are given, one with a 7-element random sample and the other with a clinical random sample of 31 participants, in which the *two-sided tie* correction does not work, but the new correction does. It is concluded that the new corrected formulas coincide with Goodman-Kruskal's gamma as compared to Glass' formula that matches Somers' d_{YX} or asymmetric measure of association of Y ranking with respect to X dichotomy. The use of this underreported coefficient is suggested, which is very easy to calculate from its equivalence with Kruskal-Wallis' gamma and Somers' d_{YX} .

Keywords

Ordinal Variable, Dichotomy, Linear Association, Nonparametric Statistics, Descriptive Statistics

1. Introduction

The objective of this statistical methodology article is to derive and verify a new correction for the rank-biserial correlation in the case of ties to achieve equality of results between Kendall's tau-type and Spearman's rho-type formulas. Its reason is that the correction given by Cureton [1] and taken up by Willson [2] does not work well. The derivation of the correction is not explicit in the articles

written by Cureton [1] and Willson [2] and is only verified by an example with a sample of 10 elements in Cureton's article [1].

It begins by defining what rank-biserial correlation is and its calculation with or without ties using Kendall's tau-type coefficient developed by Cureton [3]. The two Cureton's formulas, one to use when there are ties and the other when there are no ties, give the same result as Kruskal-Wallis' gamma, which is a non-directional association coefficient. It continues with the presentation of Spearman's rho-type formula given by Glass [4] and the one based on the Mann-Whitney U statistic [5] created by Willson [2]. These formulas yield the same result as Cureton's coefficient [3] when there are no ties. It should be noted that Glass' and Willson's formulas, whether or not there are ties, coincide with Somers' d_{YX} (of the Y ranking with respect to the X dichotomy), and Somers' three coefficients d_{YX} , $d_{X|Y}$, and d_{XY} reduce to Kruskal-Wallis' gamma when there are no ties. In this section, the asymptotic standard error formula deduced by Willson [2] from the relationship between the rank-biserial correlation and the Mann-Whitney U test is shown. This error allows us to perform a significance test and an estimation of the confidence interval. It is continued with corrections given by Cureton [1] and Willson [2] so that Spearman's rho-type and U-statistic-based formulas give the same result as Cureton's formula [3] in the case of ties. It is found that the correction does not work well, so a new correction is algebraically derived and checked with three examples. In this way, the three tie-corrected coefficients, when ties are present, coincide with each other and with Kruskal-Wallis' gamma. Also shown is the calculation of the asymptotic standard error in the case of ties and the test of significance with small samples, which Willson derived from the relationship between the rank-biserial correlation and the Mann-Whitney U test. This error and the test are valid with the new proposed correction. In the last section, the equivalence among formulas for the rank-biserial correlation is raised the other way around. Now the goal is to have the three coefficients match Somers' d_{YX} when there are ties. This is easily accomplished when the three coefficients without tie correction are used with data that have ties. In this situation, the three coefficients coincide with each other, and with Somers' d_{YX} which is a coefficient of the directional association. At this point, and returning to the third example, it is suggested that non-directionality may be more interesting than directionality of Y conditional on X. The article ends with some conclusions about the rank-biserial correlation, suggestions to promote its use, and proposals for further research.

2. Kendall's Tau-Type Formulas for Rank-Biserial Correlation

Rank-biserial correlation is a measure of association that has been developing since 1949 with Brogden's pioneering work [6]. It was specified by Cureton [3], studied and expanded by Glass [4], Stanley [7], and Willson [2], and disseminated by authors such as Khamis [8] and Berry, Johnston, and Mielke Jr. [9].

However, it is underreported [9] [10], does not appear in statistical software [11], and is not usually taught in undergraduate statistics courses [12] [13].

In 1956, the American psychologist Edward Eugene Cureton (1902-1992) developed two Kendall's tau-type formulas to measure the association between a ranking (1 to n) or a ranking variable Y and a dichotomous qualitative variable X (0 and 1) or dichotomized from an assumed or unknown ranking X' (1 to n). One formula is applied when there are ties or repeated values in Y that affect both groups of X . The other formula is used in all other cases. Cureton named this measure the rank-biserial correlation coefficient and denoted it by τ_{rb} [3]. Its potential amplitude ranges from -1 (the n_0 highest ranks of Y are associated with the n_0 values 0 of X and the n_1 lowest ranks of Y are associated with the n_1 values 1 of X) to 1 (the n_0 lowest ranks of Y are associated with the n_0 values 0 of X and the highest n_1 ranks of Y are associated with the n_1 values 1 of X). Negative values in τ_{rb} show that high ranks in Y are associated more with the value 0 than with the value 1 of X (inverse linear association). Conversely, positive values in τ_{rb} reflect that high ranks of Y are associated more with the value 1 than with the value 0 of X (direct linear association). A value of 0 in τ_{rb} shows that there is no linear association between X and Y [3].

In case there are no ties in Y (ranking), the formula for τ_{rb} is a quotient (Formula (1)). Its numerator is the difference between the number of concordant and discordant pairs. Its denominator is the product between the number of zeros and ones, that is, between the sample sizes of groups 0 and 1 of X . This coefficient coincides with the Goodman-Kruskal gamma [14], as shown in Formula (1).

$$\tau_{rb} = \frac{A - D}{n_0 \times n_1} = \frac{A - D}{A + D} = \gamma \quad (1)$$

A = agreements or number of concordant pairs with a direct association between X (dichotomy) and Y (ranking). Let be the data pairs (x_p, y_i) and (x_p, y_j) . We say that these pairs are concordant if $x_i > x_j$ and $y_i > y_j$ or $x_i < x_j$ and $y_i < y_j$.

D = disagreements or number of discordant pairs with a direct association between X (dichotomy) and Y (ranking). We say that the data pairs (x_p, y_i) and (x_p, y_j) are discordant if $x_i > x_j$ and $y_i < y_j$ or $x_i < x_j$ and $y_i > y_j$.

n_0 = sample size of group 0 of X .

n_1 = sample size of group 1 of X .

$n = n_0 + n_1$ = overall sample size.

$A + D = n_0 \times n_1$ (no ties).

Agreements (A) and disagreements (D) can be computed from a $2 \times k$ contingency table, using its joint absolute frequencies, n_{ij} . The two values of X (dichotomy) are placed per row: 0 (first row) and 1 (second row). The k values of Y (ordinal variable) are arranged per column in ascending order to the right (1 to k), as shown in **Table 1**. The sum of products between each of the first $k - 1$ frequencies in the first row (n_{11} to $n_{1,k-1}$) and the sum of the remaining frequencies to the right and below (in the second row) after removing the row and

Table 1. Joint frequencies between Y (dichotomy) per row and X (ranking or ordinal variable) per column and marginal frequencies of Y.

Y	X					n_i
	1		i		k	
0	n_{01}	...	n_{0i}	...	n_{0k}	n_0
1	n_{11}	...	n_{1i}	...	n_{1k}	n_1

column of the frequency gives the agreements or number of concordant pairs (Formula (2)).

$$A = \sum_{j=1}^{k-1} (n_{1j} \times \sum_{j'>j}^k n_{2j'}) \tag{2}$$

The sum of products between each of the last $k - 1$ frequencies in the first row (n_{12} to n_{1k}) and the sum of the remaining frequencies to the left and below (in the second row) after removing the row and column of frequency gives the disagreements or number of discordant pairs (Formula (3)).

$$D = \sum_{j=2}^k (n_{1j} \times \sum_{j'<j}^{k-1} n_{2j'}) \tag{3}$$

From **Table 1**, n_0 is the marginal frequency of the value 0 of X and is obtained through the sum of the k joint frequencies of the first row. In turn, n_1 is the marginal frequency of the value 1 of X and is obtained through the sum of the k joint frequencies of the second row (Formula (4)).

$$n_i = \sum_{j=1}^k n_{ij}; i = 0, 1 \tag{4}$$

If there are ties or repeated values in Y that affect both groups of X (0 and 1), which Cureton [3] calls *bracket ties*, a correction in the denominator of coefficient τ_{rb} is required. This corrected formula also matches the Goodman-Kruskal gamma [14], as shown in Formula (5).

$$\tau_{rb} = \frac{A - D}{n_0 \times n_1 - \sum_{l=1}^c n_{0l} \times n_{1l}} = \frac{A - D}{A + D} = \gamma \tag{5}$$

Agreements (A) and disagreements (D) are calculated from the $2 \times k$ contingency table as seen in the previous paragraphs (Formulas (2) and (3)). From **Table 1**, identifying the *bracket ties* is easy. They are those values of Y that have non-zero frequencies in both rows. The frequency in group 0 (n_{0l}) and the frequency in group 1 (n_{1l}) of each of the c *bracket ties* (or Y values that repeat in both groups 0 and 1 of X) are counted. The sum of products between both frequencies provides the tie correction (subtrahend in Formula (6)). The fact that $\tau_{rb} = \gamma$ implies that this correction made to the product between the sample sizes of the two groups of X is equal to the sum of agreements and disagreements when such ties are present (Formula (6)). In turn, it implies that the product between the sample sizes is the sum of agreement, disagreement, and tie correction.

With tied data

$$n_0 \times n_1 - \sum_{l=1}^c n_{0l} \times n_{1l} = A + D$$

$$n_0 \times n_1 = A + D + \sum_{l=1}^c n_{0l} \times n_{1l} \tag{6}$$

See an example from Cureton [1] [3] with a random sample of ten paired data (x_p, y_p) . The dichotomous qualitative variable X or dichotomized from a ranking X' has two values (0 and 1) and the ordinal variable Y has six ordered categories (1, 2, 3, 4, 5, and 6). The 12 joint frequencies of X and Y are arranged in a 2×6 contingency table (Table 2) that allows the calculation of agreements (A) and disagreements (D), the identification of *bracket ties*, and the computation of the tie correction.

Agreements or number of concordant pairs according to Formula (2).

$$A = 1 \times (0 + 2 + 1 + 1 + 2) + 2 \times (2 + 1 + 1 + 2) + 0 \times (1 + 1 + 2) + 1 \times (1 + 2) + 0 \times 2$$

$$= 6 + 12 + 0 + 3 + 0 = 21$$

Disagreements or number of discordant pairs according to Formula (3).

$$D = 0 \times (0 + 0 + 2 + 1 + 1) + 0 \times (0 + 0 + 2 + 1) + 1 \times (0 + 0 + 2)$$

$$+ 0 \times (0 + 0) + 2 \times 0$$

$$= 0 + 0 + 2 + 0 + 0 = 2$$

Bracket ties or Y values that are repeated in the two groups of X: value 4 of Y.

Tie correction or sum of products between the frequencies of the *bracket ties* in groups 0 and 1 de X according to the subtrahend of Formula (6).

$$\sum_{l=1}^1 n_{0l} \times n_{1l} = n_0 \times n_1 = 1 \times 1 = 1$$

Biserial-rank correlation coefficient according to Formula (5).

$$\tau_{rb} = \frac{A - D}{n_0 \times n_1 - \sum_{l=1}^c n_{0l} \times n_{1l}} = \frac{21 - 2}{4 \times 6 - 1} = \frac{19}{23} \approx 0.8261$$

The statement in Formula (5) that the value of the coefficient τ_{rb} [3] is equal to the Goodman-Kruskal gamma [14] is verified.

$$\gamma = \frac{A - D}{A + D} = \frac{21 - 2}{21 + 2} = \frac{19}{23} \approx 0.8261 = \tau_{rb}$$

3. Spearman's Rho-Type and U-Statistic-Based Formulas for Rank-Biserial Correlation with Untied Data

The American statistician Gene V. Glass provided three formulas corresponding

Table 2. Joint frequencies between X (dichotomy) per row and Y (ordinal variable) per column and marginal frequencies of X.

X	Y						n_i
	1	2	3	(4)	5	6	
0	1	2	0	1	0	0	4
1	0	0	2	1	1	2	6

Note. *Bracket ties* or values of Y that are repeated in both groups of X are indicated in parentheses, $n_i = \sum_j n_{ij}$ = marginal frequency of the value i of X or sum of the six joint frequencies per row. Source: Prepared by the author with data from Cureton [1] [3].

to Spearman’s rho-type coefficient for calculating rank-biserial correlation [4]. All three formulas are equivalent (Formula (7)). In addition, Glass demonstrated that these formulas are equal to Kendall’s tau-type coefficient developed by Cureton [3] when there are no ties. The new coefficient was denoted by r_{rb} and has a range from -1 to 1 .

$$\begin{aligned}
 r_{rb} &= \frac{2(\bar{r}_{Y|X=1} - \bar{r}_{Y|X=0})}{n} = \frac{2}{n_0} \times (\bar{r}_{Y|X=1} - \bar{r}_Y) = \frac{2}{n_0} \times \left(\bar{r}_{Y|X=1} - \frac{n+1}{2} \right) \\
 &= \frac{2}{n_1} \times (\bar{r}_Y - \bar{r}_{Y|X=0}) = \frac{2}{n_1} \times \left(\frac{n+1}{2} - \bar{r}_{Y|X=0} \right)
 \end{aligned}
 \tag{7}$$

n = sample size that includes n_0 values 0 and n_1 values 1 of X.

$\bar{r}_{Y|X=0}$ = conditional mean rank of Y given the value 0 of X.

$\bar{r}_{Y|X=1}$ = conditional mean rank of Y given the value 1 of X.

$\bar{r}_Y = (n + 1)/2$ = (unconditional) mean rank of Y.

Newson [15] [16] proved that the Glass coefficient r_{br} matches Somers’ [17] $d_{Y|X}$, which is an asymmetric measure of association of the ranking (variable Y) with respect to the dichotomy (X). This equality occurs whether the coefficients are computed with untied or tied data (treated by the averaged rank method). Somers’ $d_{Y|X}$ is obtained through a quotient (Formula (8)). Its numerator is the difference between agreements (A) and disagreements (D) calculated with Formulas (2) and (3). Its denominator is the sum of the agreements, disagreements and ties in Y. These ties, denoted by E_Y , are obtained by the sum of products between the frequency per column of **Table 1** (Formula (9)).

$$r_{rb} = d_{Y|X} = \frac{A - D}{A + D + E_Y}
 \tag{8}$$

$$E_Y = \sum_{j=1}^k n_{0j} \times n_{1j}
 \tag{9}$$

Kendall’s tau-type coefficient developed by Cureton [3] and Spearman’s rho-type coefficient created by Glass [4] give the same result when there are no ties or repeated values in Y (ranking) in both groups of X (dichotomy). The latter formula can be related to the Mann-Whitney U test [5], which makes it possible to calculate the rank-biserial correlation from the U-statistic, as a coefficient in absolute value or measure of effect size [2] [4]. In turn, this relationship allows obtaining a formula to compute the asymptotic standard error (ASE), based on the convergence to a normal distribution. This error makes interval estimation and significance testing of the rank-biserial correlation coefficient possible (Formula (10)). A sample of at least 20 elements in one of the two independent groups and no less than 8 in the other group is recommended to use this asymptotic approximation [2].

$$|r_{rb}| = 1 - \frac{2U}{n_0 \times n_1} \in [0,1]
 \tag{10}$$

U = Mann-Whitney U statistic [5]. The Y ranks are separated by the two groups of X, and the ranks of each group are summed (Formula (11)). These

sums of ranks allow obtaining the U_0 and U_1 statistics and the minimum of both yields the Mann-Whitney U statistic (Formula (12)).

$$\begin{aligned}
 SR_{Y|X=0} &= \sum_{i=1}^{n_0} r_{y_{i0}} \\
 SR_{Y|X=1} &= \sum_{i=1}^{n_1} r_{y_{i1}} \\
 U_0 &= n_0 n_1 + \frac{n_0(n_0+1)}{2} - SR_{Y|X=0} \\
 U_1 &= n_0 n_1 + \frac{n_1(n_1+1)}{2} - SR_{Y|X=1} \\
 \min(U_0, U_1) &= U
 \end{aligned}
 \tag{11}$$

The asymptotic standard error of r_{rb} is computed using Formula (13).

$$ASE_{r_{rb}} = \sqrt{\frac{n+1}{3n_0n_1}}
 \tag{13}$$

Confidence interval for r_{rb} is shown in Formula (14). If 0 is not included within the interval, it indicates that the population correlation is not null with a confidence level of $1 - \alpha$. Conversely, if it is included, the population correlation is not significant.

$$\begin{aligned}
 P\left(r_{rb} - z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{n+1}{3n_0n_1}} \leq \rho_{rb} \leq r_{rb} + z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{n+1}{3n_0n_1}}\right) &= 1 - \alpha \\
 \rho_{rb} \neq 0 \rightarrow 0 \notin \left[r_{rb} - z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{n+1}{3n_0n_1}}, r_{rb} + z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{n+1}{3n_0n_1}}\right]
 \end{aligned}
 \tag{14}$$

$z_{1-\alpha/2} = 1 - \alpha/2$ quantile in a standard normal distribution $N(0, 1)$. If $\alpha = 0.05$ (conventional value), $z_{0.975} = 1.96$.

4. U-Statistic-Based and Spearman’s Rho-Type Formulas for Rank-Biserial Correlation with Tied Data

In 1968, Cureton [1] remarked that Spearman’s rho-type coefficient developed by Glass [4] and Kendall’s tau-type coefficient developed by him in 1956 only coincide when there are no ties in Y affecting both groups of X. In 1956, such a tie was named a *bracket tie* [3]. In the 1968 paper, Cureton [1] proposed a formula that allows the convergence of results of the two types of coefficients when there is a tie that affects the lowest n_0 ranks and the highest n_1 ranks of Y (Formula (15)). This type of tie can be called *two-sided tie* to differentiate it from the *bracket ties* or values of Y repeated in the two groups of X. In this tie correction, the *bracket ties* are ignored and only the *two-sided tie* is used [1].

$$r_{rb} = \frac{\bar{r}_{Y|X=1} - \frac{n+1}{2}}{\frac{n_0}{2} - \frac{n_1}{n_1}}
 \tag{15}$$

In turn, when the rank-biserial correlation is calculated from the Mann-Whitney U statistic [5], Willson [2] raises the following formula for the correction in the case of a *two-sided tie* (Formula (16)).

$$|r_{rb}| = \frac{n_0 n_1}{n_0 n_1 - 2b} \times \left(1 - \frac{2U}{n_0 n_1} \right) \tag{16}$$

To obtain the value of b , the n values of Y are sorted in descending order. Unaveraged ranks are assigned to Y values in one column and averaged ranks in case of ties in another column. Ranks are separated by a horizontal line. Above the line are the n_j highest ranks, and below the line are the n_0 lowest ranks, where n_0 is the number of zeros and n_j is the number of ones in X . A *two-sided tie* is present when the same averaged rank appears above and below the line. The value of b is obtained through the difference between the sum of the unaveraged ranks of the *two-sided tie* among the n_j highest ranks and the corresponding sum of the averaged ranks [1]. If there is no *two-sided tie*, both rank sums are the same and b is 0. If b is null, the formulas are the same as those used with untied data [2].

Cureton [3] takes up the example of 10 elements, which he had previously presented in 1956 and in which a *two-sided tie* appears. With the data, he computed the described correction and checked that r_{br} gives the same result as τ_{br} (Table 3).

Table 3. Random sample of 10 elements sorted in descending order from Y -ranks.

i	x_i	y_i	r_{yi}^*	r_{yi}	r_{yi0}	r_{yi1}
9	1	6	10	9.5		9.5
10	1	6	9	9.5		9.5
8	1	5	8	8		8
6	0	4	7	6.5	6.5	
7	1	4	6	6.5		6.5
4	1	3	5	4.5		4.5
5	1	3	4	4.5		4.5
2	0	2	2	2.5	2.5	
3	0	2	3	2.5	2.5	
1	0	1	1	1	1	
n_j					4	5
$SR_j = \sum_i$					12.5	42.5
SR/n_j					3.125	7.083

Note. i = sampling order, x_i = group membership of element i = {0 = group of lowest ranks, 1 = group of highest ranks}, y_i = score in Y of element i , r_{yi}^* = ranks of Y with unaveraged ranks in case of ties, r_{yi} = ranks of Y with averaged ranks in case of ties, r_{yi0} = ranks of Y corresponding to group 0 of X , and r_{yi1} = ranks of Y corresponding to group 1 of X . Statistics: n_j = number of elements in each group, $SR_j = \sum_i$ = sum of ranks per group (per column), SR/n_j = average of the ranks per group. The two-sided tie is highlighted in bold type and its ranks above the line are in italics. Source: Prepared by the author with data from Cureton [1] [3].

$$\bar{r}_{Y|X=1} = \frac{\sum_{i=1}^{n_1} r_{yi}}{n_1} = \frac{4.5+4.5+6.5+8+9.5+9.5}{6} = 7.08\bar{3}$$

$$\bar{r}_y = \frac{n+1}{2} = \frac{n_0+n_1+1}{2} = \frac{4+6+1}{2} = \frac{11}{2} = 5.5$$

In **Table 3**, when the highest six ranks are separated from the lowest four ranks by a horizontal line, a *two-sided tie* is found that corresponds to a Y value of 3. The sum of the unaveraged ranks of the *two-sided tie* among the four ranks highest is 5 and the corresponding sum of averaged ranks is 4.5, so the value of *b* or the difference between both sums is 0.5. When Spearman’s type-rho coefficient r_{rb} is computed using Formula (15) given by Cureton [1], it gives the same result as Kendall’s tau-type coefficient (Formula (5)) developed by Cureton [3].

$$b = 5 - 4.5 = 0.5$$

$$r_{rb} = \frac{\bar{r}_{X|Y=1} - \frac{n+1}{2}}{\frac{n_0}{2} - \frac{b}{n_1}} = \frac{7.08\bar{3} - 5.5}{\frac{4}{2} - \frac{0.5}{6}} = \frac{1.58\bar{3}}{1.91\bar{6}} \approx 0.8261 = \tau_{rb}$$

$$r_{rb} = \tau_{rb} = \frac{A - D}{n_0 \times n_1 - \sum_{l=1}^c n_{0l} \times n_{1l}} = \frac{19}{23} \approx 0.8261$$

From the formula based on the U-statistic to calculate r_{rb} [2], equality of results with τ_{rb} (Formula (5)) is also achieved. The sums of ranks $SR_{Y|X=0}$ and $SR_{Y|X=1}$ are calculated using Formula (11) and the U_0 and U_1 statistics using Formula (12). The minimum of the latter two is the U-statistic (Formula (12)). Applying Formula (16), the rank biserial correlation is obtained.

$$SR_{Y|X=0} = \sum_{i=1}^4 r_{yi} = 1 + 2.5 + 2.5 + 6.5 = 12.5$$

$$SR_{Y|X=1} = \sum_{i=1}^6 r_{yi} = 4.5 + 4.5 + 6.5 + 8 + 9.5 + 9.5 = 42.5$$

$$U_0 = n_0 n_1 + \frac{n_0(n_0+1)}{2} - SR_{r_y|x=0} = 4 \times 6 + \frac{4 \times 5}{2} - 12.5 = 21.5$$

$$U_1 = n_0 n_1 + \frac{n_1(n_1+1)}{2} - SR_{r_y|x=1} = 4 \times 6 + \frac{6 \times 7}{2} - 42.5 = 2.5$$

$$U = \min(U_0, U_1) = \min(21.5, 2.5) = 2.5$$

$$|r_{rb}| = \frac{n_0 n_1}{n_0 n_1 - 2b} \times \left(1 - \frac{2U}{n_0 n_1} \right) = \frac{4 \times 6}{4 \times 6 - 2 \times 0.5} \times \left(1 - \frac{2 \times 2.5}{4 \times 6} \right)$$

$$= \frac{24}{23} \times \left(1 - \frac{5}{24} \right) = \frac{24}{23} \times 0.791\bar{6} = |\tau_{rb}|$$

$$= \frac{|A - D|}{n_0 n_1 - \sum_{l=1}^c n_{0l} \times n_{1l}} = \frac{19}{23} \approx 0.8261$$

It should be noted that Formula (15) given by Cureton [1] does not work well in all cases with a *two-sided tie*. In turn, the other formula based on the Mann-Whitney U statistic (Formula (16)) developed by Willson [2] also does not fit this definition of *b*. However, both formulas do achieve equality of results with

Kendall’s tau-type formula when using the correction based on *bracket ties* given by Cureton [3] for τ_{rb} . The constant b is half of the sum of products between the frequencies in groups 0 and 1 of X of the c *bracket ties*, $b = (\sum n_{0l} \times n_{1l})/2$, in Cureton’s formula [1]. In Willson’s formula [2], the correction is the full value of the sum of products, $2 \times b = \sum n_{0l} \times n_{1l}$. To demonstrate algebraically why this correction works (Formula (17)), we start from an equality given by Glass [4] in his proof that $r_{rb} = \tau_{rb}$ when there are no ties.

$$A - D = 2 \left[\left(\sum_{i=1}^{n_1} r_{y_i|x=1} \right) - \frac{n_1(n+1)}{2} \right] = 2 \left(n_1 \bar{r}_{y|x=1} - \frac{n_1(n+1)}{2} \right)$$

Both sides of the equality are divided by $n_0 n_1 - \sum_{l=1}^c n_{0l} \times n_{1l}$

$$\frac{A - D}{n_0 n_1 - \sum_{l=1}^c n_{0l} \times n_{1l}} = \frac{2 \left(n_1 \bar{r}_{y|x=1} - \frac{n_1(n+1)}{2} \right)}{n_0 n_1 - \sum_{l=1}^c n_{0l} \times n_{1l}}$$

$$\tau_{rb} = \frac{2 \left(n_1 \bar{r}_{y|x=1} - \frac{n_1(n+1)}{2} \right)}{n_0 n_1 - \sum_{l=1}^c n_{0l} \times n_{1l}}$$

The numerator and denominator are divided by the inverse of n_1 .

$$\begin{aligned} & \frac{2 \left(n_1 \bar{r}_{y|x=1} - \frac{n_1(n+1)}{2} \right)}{n_1 \left(n_0 n_1 - \sum_{l=1}^c n_{0l} \times n_{1l} \right)} = \frac{2 \left(\frac{n_1 \bar{r}_{y|x=1}}{n_1} - \frac{n_1(n+1)}{2n_1} \right)}{\frac{n_0 n_1}{n_1} - \frac{\sum_{l=1}^c n_{0l} \times n_{1l}}{n_1}} = \frac{2 \left(\bar{r}_{y|x=1} - \frac{n+1}{2} \right)}{n_0 - \frac{\sum_{l=1}^c n_{0l} \times n_{1l}}{n_1}} \\ & = \frac{\bar{r}_{y|x=1} - \frac{n+1}{2}}{\frac{1}{2} \left(n_0 - \frac{\sum_{l=1}^c n_{0l} \times n_{1l}}{n_1} \right)} = \frac{\bar{r}_{y|x=1} - \frac{n+1}{2}}{\frac{n_0}{2} - \frac{\sum_{l=1}^c n_{0l} \times n_{1l}}{2n_1}} \end{aligned}$$

The new tie correction, which is denoted by b^* , is half of the sum of products between the frequencies in groups 0 and 1 of X of the c *bracket ties* (Formula (17)).

$$b^* = \frac{\sum_{l=1}^c n_{0l} \times n_{1l}}{2} \tag{17}$$

The correction b^* is applied to the formula given by Cureton [3] without requiring any changes (Formula (18)) and gives the same result as its tau-type coefficient for rank-biserial correlation (Formula (5)).

$$r_{rb}^* = \frac{\bar{r}_{y|x=1} - \frac{n+1}{2}}{\frac{n_0}{2} - \frac{b^*}{n_1}} = \tau_{rb} \tag{18}$$

It is also applied to the formula given by Willson [2] without requiring any changes (Formula (19)) and gives the same result as Cureton [1] tau-type coefficient (Formula (5)).

$$|r_{rb}^*| = \frac{n_0 n_1}{n_0 n_1 - \sum_{l=1}^c n_{0l} \times n_{1l}} \times \left(1 - \frac{2U}{n_0 n_1}\right) = \frac{n_0 n_1}{n_0 n_1 - 2b^*} \times \left(1 - \frac{2U}{n_0 n_1}\right) = |\tau_{rb}| \quad (19)$$

This new proposal is verified with the previous example of 10 elements in which Formula (15) of Cureton [1] and Formula (16) of Willson [2] work well with the constant b defined from the *two-sided tie*, that is, they achieve the same result as Cureton’s tau-type coefficient (Formula (5)) and, therefore, as the Goodman-Kruskal gamma [14]. The constant b is calculated by Formula (17) and used in Formulas (18) (rho-type coefficient) and (19) (U-statistic-based coefficient).

$$b^* = \frac{\sum_{l=1}^1 n_{0l} \times n_{1l}}{2} = \frac{n_0 \times n_1}{2} = \frac{1 \times 1}{2} = \frac{1}{2} = 0.5$$

$$r_{rb}^* = \frac{\bar{r}_{Y|X=1} - \frac{n+1}{2}}{\frac{n_0}{2} - \frac{b^*}{n_1}} = \frac{7.08\bar{3} - 5.5}{\frac{4}{2} - \frac{0.5}{6}} = \frac{1.58\bar{3}}{1.91\bar{6}} = \tau_{rb} = \frac{A - D}{n_0 n_1 - \sum_{l=1}^c n_{0l} \times n_{1l}}$$

$$= \frac{19}{23} \approx 0.8261$$

$$U = \min(U_0, U_1) = \min(21.5, 2.5) = 2.5$$

$$|r_{rb}^*| = \frac{n_0 n_1}{n_0 n_1 - 2b^*} \times \left(1 - \frac{2U}{n_0 n_1}\right) = \frac{4 \times 6}{4 \times 6 - 2 \times 0.5} \times \left(1 - \frac{2 \times 2.5}{4 \times 6}\right)$$

$$= \frac{24}{23} \times \left(1 - \frac{5}{24}\right) = \frac{24}{23} \times 0.791\bar{6} = |\tau_{rb}| = \frac{19}{23} \approx 0.8261$$

A new seven-element example with two *bracket ties* and a *two-sided tie* is presented. With these data, Formula (15) of Cureton [1] and Formula (16) of Willson [2] fail to coincide with τ_{rb} (Formula (5)), when Formula (18) and (19), with the constant b^* defined from the *bracket ties*, does achieve the same results as τ_{rb} (Formula (5)). The data of the seven elements are shown in **Table 4**, which is a 2×4 contingency table between the dichotomous variable X or dichotomized from a ranking X' (0 and 1) and the ordinal variable Y (1, 2, 3, and 4). This table allows the calculation of agreements (A) and disagreements (D), the identification of *bracket ties*, and the computation of the correction based on *bracket ties*.

Table 4. Joint frequencies between X (dichotomy) per row and Y (ordinal variable) per column and marginal frequencies of X.

X	Y				n_i
	1	(2)	(3)	4	
0	1	1	1	0	3
1	0	2	1	1	4

Note. *Bracket ties* or values of Y that are repeated in both groups of X are indicated in parentheses, $n_i = \sum_j p_{ij}$ = marginal frequency of the value i X or sum of the four joint frequencies per row. Source: Prepared by the author.

We start by calculating Cureton’s tau-type coefficient [3], for which agreements or number of concordant pairs and disagreements or number of discordant pairs are computed using Formulas (2) and (3).

$$A = 1 \times (2 + 1 + 1) + 1 \times (1 + 1) + 1 \times 1 = 4 + 2 + 1 = 7$$

$$D = 0 \times (0 + 2 + 1) + 1 \times (0 + 2) + 1 \times 0 = 0 + 2 + 0 = 2$$

Bracket ties or Y values that are repeated in the two groups of X are identified: values 2 and 3 of Y.

Tie correction or sum of products between the frequencies of the *bracket ties* in groups 0 and 1 de X is obtained (subtrahend of Formula (6))

$$\sum_{i=1}^2 n_{0i} \times n_{1i} = 1 \times 2 + 1 \times 1 = 3$$

After these calculations, Kendall’s tau-type coefficient for the rank-biserial correlation can be calculated using Formula (5), which yields a value of zero point repeating five.

$$\tau_{rb} = \frac{A - D}{n_0 \times n_1 - \sum_{i=1}^c n_{0i} \times n_{1i}} = \frac{7 - 2}{3 \times 4 - 3} = \frac{5}{9} \approx 0.5$$

The statement of Formula (5) that the τ_{rb} [3] coincides with the value of the Goodman-Kruskal gamma [14] is verified.

$$\gamma = \frac{A - D}{A + D} = \frac{7 - 2}{7 + 2} = \frac{5}{9} = 0.5 = \tau_{rb}$$

In **Table 5**, seven sample data are sorted in descending order from the ranks of Y (with averaged ranks in case of ties). When the four highest ranks are separated from the three lowest by a horizontal line, a *two-sided tie* is discovered and this corresponds to a Y value of 2. The sum of the unaverage ranks of the *two-sided tie* among the four highest ranks is 4 and the corresponding sum of averaged ranks is 3, so the value of *b* or the difference of both sums is 1.

Two-sided tie: value 2 of Y.

$$b = 4 - 3 = 1$$

The corrected Spearman’s rho-type coefficient [1] is calculated according Formula (15).

$$r_{rb} = \frac{\bar{r}_{Y|X=1} - \frac{n+1}{2}}{\frac{n_0}{2} - \frac{b}{n_1}} = \frac{4.625 - 4}{\frac{3}{2} - \frac{1}{4}} = \frac{0.625}{1.25} = 0.5 \neq \tau_{rb} = 0.5$$

The corrected formula for rank-biserial correlation from U-statistic [2] is computed according Formula (16). The sums of ranks $SR_{Y|X=0}$ and $SR_{Y|X=1}$ using Formula (11) are calculated and the U_0 and U_1 statistics using Formula (12). The minimum of the latter two is the U-statistic.

$$SR_{Y|X=0} = \sum_{i=1}^3 r_{yi} = 1 + 3 + 5.5 = 9.5$$

$$SR_{Y|X=1} = \sum_{i=1}^4 r_{yi} = 3 + 3 + 5.5 + 7 = 18.5$$

Table 5. Random sample of seven elements sorted in descending order from Y-ranks.

i	x_i	y_i	r_{yi}^*	r_{yi}	r_{yio}	r_{yil}
7	1	4	7	7		7
5	0	3	6	5.5	5.5	
6	1	3	5	5.5		5.5
3	0	2	4	3	3	
2	1	2	3	3		3
4	1	2	2	3		3
1	0	1	1	1	1	
n_j					3	4
$SR_j = \sum_i$					9.5	18.5
SR/n_j					3.167	4.625

Note. i = sampling order, x_i = group membership of element i = {0 = group of lowest ranks, 1 = group of highest ranks, y_i = score in Y of element i , r_{yi}^* = ranks of Y with un-averaged ranks in case of ties, r_{yi} = ranks of Y with averaged ranks in case of ties, r_{yio} = ranks of Y corresponding to group 0 of X, and r_{yil} = ranks of Y corresponding to group 1 of X. Statistics: n_j = number of elements in each group, $SR_j = \sum_i$ = sum of ranks per group (per column), SR/n_j = average of the ranks per group. The two-sided tie is highlighted in bold type and its ranks above the line are in italics. Source: Prepared by the author.

$$U_0 = n_0n_1 + \frac{n_0(n_0 + 1)}{2} - SR_{Y|X=0} = 3 \times 4 + \frac{3 \times 4}{2} - 9.5 = 8.5$$

$$U_1 = n_0n_1 + \frac{n_1(n_1 + 1)}{2} - SR_{Y|X=1} = 3 \times 4 + \frac{4 \times 5}{2} - 18.5 = 3.5$$

$$U = \min(U_0, U_1) = \min(8.5, 3.5) = 3.5$$

$$|r_{rb}| = \frac{n_0n_1}{n_0n_1 - 2b} \times \left(1 - \frac{2U}{n_0n_1}\right) = \frac{3 \times 4}{3 \times 4 - 2 \times 1} \times \left(1 - \frac{2 \times 3.5}{3 \times 4}\right) = 1.2 \times 0.41\bar{6} = 0.5 \neq |\tau_{br}| = 0.5$$

Formula (15) of Cureton [1] and Formula (16) of Willson [2] give the same result, which is zero point five, but differ from the result of Cureton’s τ_{rb} (Formula (5)), which is zero point repeating five. However, the calculation of r_{rb} using Formulas (18) and (19), with b^* defined from the *bracket ties* (Formula (17)), does achieve the same result as Formula (5). First, the tie correction b^* is calculated using Formula (17).

$$b^* = \frac{\sum_{l=1}^2 n_{0l} \times n_{1l}}{2} = \frac{1 \times 2 + 1 \times 1}{2} = \frac{3}{2} = 1.5$$

Next, with this constant b^* , Formula (18) (rho-type coefficient) and Formula (19) (U-statistic-based coefficient) are used to compute the rank-biserial correlation, and both yield a result of zero point repeating five.

$$\begin{aligned}
 r_{rb}^* &= \frac{\bar{y}_{Y|X=1} - \frac{n+1}{2}}{\frac{n_0}{2} - \frac{b^*}{n_1}} = \frac{4.625 - 4}{\frac{3}{2} - \frac{1.5}{4}} = \frac{0.625}{1.125} = 0.5 = \tau_{rb} = \frac{A - D}{n_0 n_1 - \sum_{l=1}^c n_{0l} \times n_{1l}} \\
 &= \frac{5}{9} = 0.5 = \gamma = \frac{A + D}{A - D} = \frac{5}{9} = 0.5 \\
 |r_{rb}^*| &= \frac{n_0 n_1}{n_0 n_1 - 2b^*} \times \left(1 - \frac{2U}{n_0 n_1}\right) = \frac{3 \times 4}{3 \times 4 - 2 \times 1.5} \times \left(1 - \frac{2 \times 3.5}{3 \times 4}\right) \\
 &= \frac{12}{9} \times \left(1 - \frac{7}{12}\right) = 1.3 \times 0.41\bar{6} = 0.5 = |\tau_{rb}| = |\gamma|
 \end{aligned}$$

5. Asymptotic Standard Error with Tied Data

From the relationship between r_{rb} and the Mann-Whitney U test [5], Willson [2] derives an asymptotic standard error for the significance test (Formula (20)) and interval estimation when there are ties in the sample data for Y. This error also applies to the new definition of b^* (Formula (17)), which is the only change in the Formulas (15) and (16) to achieve the originally intended goal: $r_{rb} = \tau_{rb}$.

$$ASE_{r_{rb}} = \sqrt{\frac{n^3 - n - \sum_{l=1}^c (t_{x_l}^3 - t_{x_l})}{3n(n-1)n_0 n_1}} \tag{20}$$

t_l = non-unit frequencies of the Y values (of the c tied values) in the total sample, where $c \leq k$ and k is the number of ordered categories of Y.

$$\begin{aligned}
 P \left(r_{rb} - z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{n^3 - n - \sum_{l=1}^k (t_{x_l}^3 - t_{x_l})}{3n(n-1)n_0 n_1}} \leq \rho_{rb} \right. \\
 \left. \leq r_{rb} + z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{n^3 - n - \sum_{l=1}^k (t_{x_l}^3 - t_{x_l})}{3n(n-1)n_0 n_1}} \right) = 1 - \alpha
 \end{aligned}$$

6. Statistical Significance with Small Samples

In the case of small samples, a critical value for r_{rb} can be used to test its statistical significance [2]. The calculation of the critical value of the rank-biserial correlation (with a sample size of n_0 for the group with the smallest sum of ranks and n_1 for the group with the largest sum of ranks, as well as a significance level α in a two-tailed test or $\alpha/2$ in a one-tailed test) requires the critical value of the U statistic for these same specifications (n_0 , n_1 , α , and a one-tailed or two-tailed test). The calculation of the critical value for r_{rb} uses the formula relating the rank-biserial correlation to the Mann-Whitney U statistic [5] when there are no ties (Formula (10)), which yields a value from 0 to 1 (Formula (21)). This absolute critical value $|r_{crit}|$ is used regardless of whether r_{rb} is calculated using Formula (10) (for untied data) or Formula (16) (for tied data).

$$H_0 : \rho_{rb} = 0$$

$$H_1 : \rho_{rb} \neq 0$$

$$|r_{crit}| = 1 - 2 \times \frac{\alpha U_{n_0 n_1}}{n_0 n_1} \in [0, 1] \quad (21)$$

If $|r_{rb}| \leq r_{crit}$, H_0 is hold.

If $|r_{rb}| > r_{crit}$, H_0 is rejected.

The previous example is taken up again with its small sample of 10 elements (four in group 0 and six in group 1). If the test is two-tailed with a significance level of 0.05, the rank-biserial correlation coefficient is significant. The $|r_{rb}|$ value is greater than the critical value, so the null hypothesis of no correlation is rejected. The critical value for r_{rb} is calculated using Formula (21).

$$|r_{crit}| = 1 - 2 \times \frac{\alpha=0.05 U_{n_0=4, n_1=6}}{n_0 n_1} = 1 - 2 \times \frac{5}{4 \times 6} = 1 - 0.41\bar{6} = 0.58\bar{3}$$

$$r_{rb} = \frac{19}{23} \approx 0.8261 > |r_{crit}| = 0.58\bar{3}$$

In the second example with an even smaller sample of seven elements (three in group 0 and four in group 1), the critical value of U is not defined at a significance level of 0.05 in a two-tailed test, so the critical value for $|r_{rb}|$ cannot be obtained. However, it is defined at a significance level of 0.1 in a one-tailed test ($\alpha = 0.2$ in a two-tailed test). In this case, the value of the rank-biserial coefficient is less than that of the critical value, so the correlation is not significant. The critical value for r_{rb} is calculated using Formula (21).

$$|r_{crit}| = 1 - 2 \times \frac{\alpha/2=0.1 U_{n_0=3, n_1=4}}{n_0 n_1} = 1 - 2 \times \frac{1}{3 \times 4} = 1 - 0.1\bar{6} = 0.8\bar{3}$$

$$r_{rb} = 0.5 < |r_{crit}| = 0.8\bar{3}$$

7. Example with a Clinical Sample of 31 Participants

A third example is presented with a larger clinical sample than the previous ones. A random sample of 31 middle-aged women, 20 with diabetes mellitus and 11 without diabetes, was recruited and their socioeconomic status, $SES = \{1 = \text{low}, 2 = \text{medium-low}, 3 = \text{medium}, 4 = \text{medium-high}, \text{and } 5 = \text{high}\}$, was recorded. The objective was to find out whether the relationship between clinical diabetes mellitus status, $DM = \{0 = \text{no case}, 1 = \text{case}\}$, and socioeconomic status is significant at a significance level of 0.05 in a two-tailed test, using the rank-biserial correlation.

The data of the 31 participants are shown in **Table 6**. In this 2×5 contingency table, the dichotomous variable of health status (0 = without diabetes and 1 = with diabetes) is placed per row and the ordinal variable of socioeconomic status with five ordered categories is arranged per column. This contingency table is made to facilitate the computation of agreements (A) and disagreements (D), the identification of *bracket ties* (SES values that are repeated in both groups of DM), and the achievement of the correction based on *bracket ties*. All of them

Table 6. Joint frequencies between diabetes mellitus (dichotomy) per row and socioeconomic status (ordinal variable) per column and marginal frequencies of diabetes mellitus.

Diabetes mellitus	Socioeconomic status					n_i
	(1)	(2)	(3)	4	5	
No case	2	2	4	2	1	11
Case	6	9	5	0	0	20

Note. Diabetes mellitus (DM): 0 = no case and 1 = case, socioeconomic status (SES): 1 = low, 2 = medium-low, 3 = medium, 4 = medium-high, and 5 = high. *Bracket ties* are identified with the ordered categories of SES that are repeated in the two DM groups and are placed in parentheses, $n_i = \sum_j n_{ij}$ = marginal frequency of the value i in DM or the sum of the five joint frequencies per row. Source: elaborated by the author.

are calculations are required by Kendall’s tau-type coefficient given by Cureton [3] to estimate the rank-biserial correlation (Formula (5)).

The sum of products between each of the first four frequencies (1 to 4) in the first row and the sum of the remaining frequencies to the right and below (in the second row) after removing the row and column of the frequency provides the agreements or concordant pairs (Formula (2)).

$$A = 2 \times (9 + 5 + 0 + 0) + 2 \times (5 + 0 + 0) + 4 \times (0 + 0) + 2 \times 0 = 28 + 10 + 0 + 0 = 38$$

The sum of products between each of the last four frequencies (from k to 2) in the first row and the sum of the remaining frequencies to the left and below (in the second row) after removing the row and column of the frequency provides the disagreements or discordant pairs (Formula (3)).

$$D = 1 \times (6 + 9 + 5 + 0) + 2 \times (6 + 9 + 5) + 4 \times (6 + 9) + 2 \times 6 = 20 + 40 + 60 + 12 = 132$$

The *bracket ties* in this example are the SES values 1, 2, and 3. The sum of products between the frequencies in groups 0 and 1 of DM (dichotomy) of the *bracket ties* in SES (ordinal variable) provides the correction that appears in the denominator of Formula (5) as a subtrahend.

$$\sum_{l=1}^c n_{0l} \times n_{1l} = 2 \times 6 + 2 \times 9 + 4 \times 5 = 12 + 18 + 20 = 50$$

The coefficient is calculated following Formula (5), as there are *bracket ties*.

$$\tau_{rb} = \frac{A - D}{n_0 \times n_1 - \sum_{l=1}^c n_{0l} \times n_{1l}} = \frac{38 - 132}{11 \times 20 - 50} = \frac{-94}{170} = -\frac{47}{85} \approx -0.5529$$

It is found that the rank-biserial correlation, estimated by Kendall’s tau-type coefficient given by Cureton [3], yields the same result as the Goodman-Kruskal gamma [14], as stated by Formula (5), and this value corresponds to a large strength of association, that is, greater than 0.50 and less than 0.70 [18].

$$\gamma = \frac{A - D}{A + D} = \frac{38 - 132}{38 + 132} = \frac{-94}{170} = -\frac{47}{85} \approx -0.5529$$

$$\tau_{rb} = \gamma$$

In **Table 7**, the data from the 31 participants are ranked. Unaveraged ranks appear in one column and averaged ranks (in case of ties) in another. The data are sorted in descending order from the Y ranks. The 20 highest ranks are separated from the 11 lowest by a horizontal line, thus uncovering a *two-way tie* corresponding to SES value of 2 (medium-low). The sum of the unaveraged ranks of the *two-way tie* among the highest ranks is 124 and the corresponding sum of averaged ranks is 112, so the value of b or the difference between the two sums is 12.

Two-sided tie: value 2 (medium-low) of SES

$$b = 124 - 112 = 12$$

The corrected Spearman's rho-type coefficient is calculated according to Formula (15).

$$r_{rb} = \frac{\bar{r}_{SES|DM=1} - \frac{n+1}{2}}{\frac{n_0}{2} - \frac{b}{n_1}} = \frac{\frac{273}{20} - \frac{31+1}{2}}{\frac{11}{2} - \frac{12}{20}} = \frac{-2.35}{4.9} \approx -0.4796 \neq \tau_{rb} = -\frac{47}{85} \approx -0.5529$$

The corrected formula for rank-biserial correlation from U-statistic is computed using Formula (16). The sums of ranks $SR_{SES|DM=0}$ and $SR_{SES|DM=1}$ are calculated with Formula (11) and the U_0 and U_1 statistics with Formula (12). The minimum of the latter two is the U-statistic.

$$SR_{SES|DM=0} = \sum_{i=1}^{11} r_{yi} = 31 + 2 \times 29.5 + 4 \times 24 + 2 \times 14 + 2 \times 4.5 = 223$$

$$SR_{SES|DM=1} = \sum_{i=1}^{20} r_{yi} = 5 \times 24 + 9 \times 14 + 6 \times 4.5 = 273$$

$$U_0 = n_0 n_1 + \frac{n_0(n_0+1)}{2} - SR_{SES|DM=0} = 11 \times 20 + \frac{11 \times 12}{2} - 223 = 63$$

$$U_1 = n_0 n_1 + \frac{n_1(n_1+1)}{2} - SR_{SES|DM=1} = 11 \times 20 + \frac{20 \times 21}{2} - 273 = 157$$

$$U = \min(U_0, U_1) = \min(63, 157) = 63$$

$$\begin{aligned} |r_{rb}| &= \frac{n_0 n_1}{n_0 n_1 - 2b} \times \left(1 - \frac{2U}{n_0 n_1}\right) = \frac{11 \times 20}{11 \times 20 - 2 \times 12} \times \left(1 - \frac{2 \times 63}{11 \times 20}\right) \\ &= \frac{220}{196} \times 0.427 = \frac{55}{49} \times 0.427 \approx 0.4796 \neq |\tau_{rb}| = -\frac{2.35}{4.9} \approx 0.5529 \end{aligned}$$

It is important to clarify that Glass [4], in his proofs, uses an operational assignment of the values of the dichotomous variable of membership: {0 = group with the lowest sum of ranks or lowest ranks, 1 = group with the highest sum of ranks or higher ranks}. However, this classification can be reversed, since the rank-biserial correlation coefficient will be positive or negative, but with the same absolute value, depending on the assignment of the labels 0 and 1 to two values of Y. In clinics, 0 is often used for no cases and 1 for cases of a disease or disorder, as in this example, although these labels are ultimately arbitrary.

The rank-biserial correlation is recalculated with Spearman's rho-type coefficient [1], using the tie correction based on *bracket ties* (Formula (17)). Applying

Table 7. Random sample of 31 participants sorted in descending order from the SES ranks, and SES marginal frequencies.

i	DM_i	SES_i	r_{SES}	r_{SES}^*	$r_{SES/DM=0}$	$r_{SES/DM=1}$	SES	t	$t^2 - t$
8	0	5	31	31	31		1	8	504
7	0	4	30	29.5	29.5		2	11	1320
12	0	4	29	29.5	29.5		3	9	720
17	0	3	28	24	24		4	2	6
19	0	3	27	24	24		5	1	0
20	0	3	26	24	24				
25	0	3	25	24	24				
6	1	3	24	24		24			
13	1	3	23	24		24			
24	1	3	22	24		24			
27	1	3	21	24		24			
29	1	3	20	24		24			
14	0	2	19	14	14				
16	0	2	18	14	14				
2	1	2	17	14		14			
9	1	2	16	14		14			
15	1	2	15	14		14			
18	1	2	14	14		14			
21	1	2	13	14		14			
22	1	2	12	14		14			
23	1	2	11	14		14			
30	1	2	10	14		14			
31	1	2	9	14		14			
4	0	1	8	4.5	4.5				
28	0	1	7	4.5	4.5				
1	1	1	6	4.5		4.5			
3	1	1	5	4.5		4.5			
5	1	1	4	4.5		4.5			
10	1	1	3	4.5		4.5			
11	1	1	2	4.5		4.5			
26	1	1	1	4.5		4.5			
n_j					11	20			

Continued

Σ_i	223	273	31	9300
Σ/n_j	20.273	13.65		

Note. i = sampling order, DM_i = group membership to participant i in $DM = \{0 = \text{no case of diabetes mellitus}, 1 = \text{case of diabetes mellitus}\}$, SES_i = ordered category of participant i in socioeconomic status (SES), r_{SES} = ranks of socioeconomic status with unaveraged ranks in cases of ties, r_{SES}^* = ranks of socioeconomic status with averaged ranks in case of ties, $r_{SES|DM=0}$ = conditional ranks of socioeconomic status given the value 0 (no case) in DM (diabetes mellitus) and $r_{SES|DM=1}$ = conditional ranks of socioeconomic status given the value 1 (case) in DM (diabetes mellitus), n_j = number of participants in each group of X, $SR_j = \Sigma_i = \text{sum per column}$, $\Sigma/n_j = \text{average of the ranks per group}$, SES = ordinated categories of socioeconomic status (1 = low, 2 = medium-low, 3 = medium, 4 = medium-high, and 5 = high), t_y = absolute frequency or times that each ordered category of SES is repeated in the total sample, $t_y^3 - t_y$ = difference of the frequency raised to cube and the frequency of each ordered category of the socioeconomic status. The two-sided tie is highlighted in bold type and its ranks above the line are in italics. Source: elaborated by the author.

this correction according to Formula (18), the results are the same as Kendall's tau-type coefficient developed by Cureton [3], as was the goal of Cureton [1] and Willson [2].

b^* = half of the sum of products between the frequencies in groups 0 (no case) and 1 (case) of DM in the *c bracket ties* of SES (Formula (17)).

$$b^* = \frac{\sum_{l=1}^3 n_{0l} \times n_{1l}}{2} = \frac{2 \times 6 + 2 \times 9 + 4 \times 5}{2} = \frac{50}{2} = 25$$

$$r_{rb}^* = \frac{\bar{r}_{SES|DM=1} - \frac{n+1}{2}}{\frac{n_0}{2} - \frac{b^*}{n_1}} = \frac{13.65 - 16}{\frac{11}{2} - \frac{25}{20}} = \frac{-2.35}{4.25} = \tau_{rb} = \gamma = -\frac{47}{85} \approx -0.5529$$

The result is the same, except that in absolute value, if the Willson's formula with the Mann-Whitney U statistic [5] is used to estimate the rank-biserial correlation (Formula (19)), but applying the *b*-correction based on *bracket ties* instead of a *two-sided tie*.

$$U_0 = n_0 n_1 + \frac{n_0(n_0 + 1)}{2} - SR_{SES|DM=0} = 11 \times 20 + \frac{11 \times 12}{2} - 223 = 63$$

$$U_1 = n_0 n_1 + \frac{n_1(n_1 + 1)}{2} - SR_{SES|DM=1} = 11 \times 20 + \frac{20 \times 21}{2} - 273 = 157$$

$$U = \min(U_0, U_1) = 63$$

$$|r_{rb}^*| = \frac{n_0 n_1}{n_0 n_1 - 2b^*} \times \left(1 - \frac{2U}{n_0 n_1}\right) = \frac{11 \times 20}{11 \times 20 - 2 \times 25} \times \left(1 - \frac{2 \times 63}{11 \times 20}\right)$$

$$= \frac{220}{170} \times \left(1 - \frac{126}{220}\right) = \frac{22}{17} \times 0.427 = |\tau_{rb}| = |\gamma| \approx 0.5529$$

Once the biserial-rank correlation has been pointwise estimated, its asymp-

otic standard error (ASE) is calculated using Formula (20). This approximation is adequate, since the group with diabetes ($DM = 1$) has 20 participants and the group without diabetes ($DM = 0$) counts more than 8 ($n_0 = 11$), giving a total sample of 31 participants [2]. See **Table 7** for the SES marginal frequencies used to calculate the tie correction in the denominator of Formula (20).

$$ASE_{r_{rb}} = \sqrt{\frac{n^3 - n - \sum_{l=1}^k (t_{x_l}^3 - t_{x_l})}{3n(n-1)n_0n_1}} = \sqrt{\frac{31^3 - 31 - 2550}{3 \times 31 \times 30 \times 11 \times 20}}$$

$$= \sqrt{0.0443304} \approx 0.2105$$

The statistical significance of the rank-biserial correlation coefficient is tested. The statistical hypothesis is two-tailed.

$$H_0 : \rho_{rb} = 0$$

$$H_1 : \rho_{rb} \neq 0$$

The z-statistic in absolute value is greater than the critical value and its probability value is less than the significance level, so the null hypothesis of zero correlation at a significance level of 0.05 in a two-tailed test is rejected.

$$|Z| = \frac{|r_{rb}|}{ASE_{r_{rb}}} = \frac{0.5529}{0.2105} = 2.6262 > z_{0.975} = 1.96$$

$$sig. = 2 \times (1 - P(Z \leq |z|)) = 2 \times (1 - P(Z \leq 2.6262)) = 0.0043 < \alpha = 0.05$$

The rank-biserial coefficient is estimated by the interval with a 95% confidence level. This confidence interval does not include 0, since the correlation is significant.

$$P\left(r_{rb} - z_{1-\frac{\alpha}{2}} \times ASE_{r_{rb}} \leq \rho_{rb} \leq r_{rb} + z_{1-\frac{\alpha}{2}} \times ASE_{r_{rb}}\right) = 1 - \alpha$$

$$P(-0.5529 - 1.96 \times 0.2105 \leq \rho_{rb} \leq -0.5529 + 1.96 \times 0.2105) = 0.95$$

$$P(\rho_{rb} \in [-0.9656, -0.1403]) = 0.95$$

It should be noted that when the rank-biserial correlation is calculated using the Spearman's rho-type coefficient given by Glass (Formula (7)) with current data, having SES ties in both DM groups, this coefficient yields the same result as the Somers [17] asymmetric d coefficient of the ordinal variable with respect to the dichotomous variable ($d_{SES|DM}$) and, therefore, differs from the result of the Goodman-Kruskal gamma [14].

$$r_{rb} = \frac{2(\bar{r}_{SES|DM=1} - \bar{r}_{SES|DM=0})}{n} = \frac{2 \times \left(\frac{273}{20} - \frac{223}{11}\right)}{31} = \frac{2 \times \left(\frac{273}{20} - \frac{223}{11}\right)}{31}$$

$$= \frac{2 \times (13.65 - 20.27)}{31} = 0.427$$

For the computation of the Somers [17] asymmetric measure of the ordinal variable (SES) with respect to the dichotomous variable (DM), data from **Table**

6 are used following the Formula (8). Apart from agreements (Formula (2)) and disagreements (Formula (3)), it is necessary to calculate ties per column or non-concordant and non-discordant pairs that are tied in SES. They are obtained through the sum of products between the two frequencies in each column of the 2×5 contingency table (Formula (9) using data of **Table 6**). This table has the two values of DM (0 = no case and 1 = case) per row and the five ordered categories of SES per column.

$$d_{SES|DM} = \frac{A - D}{A + D + E_{SES}} = \frac{38 - 132}{38 + 132 + 50} = \frac{-94}{220} = -0.427$$

$$E_{SES} = \sum_{j=1}^k n_{0j} \times n_{1j} = 2 \times 6 + 2 \times 9 + 4 \times 5 + 2 \times 0 + 1 \times 0 = 50 = \sum_{l=1}^c n_{0l} \times n_{1l}$$

$$r_{rb} = -0.427 = d_{SES|DM} \neq \gamma = -\frac{47}{85} \approx -0.5529$$

8. The Other Way Around: Tau-Type, Rho-Type, and U-Statistic-Based Formulas for Rank-Biserial Correlation Equivalent to Somer's $d_{Y|X}$ (Ranking with Respect to Dichotomy)

If any frequency per column is null in **Table 6**, the summand corresponding to that socioeconomic status category (untied value) is canceled within the sum of products in the computation ties per column, so E_{SES} equals the sum of products between the frequencies of the c bracket ties in groups 0 and 1 of DM. From Formula (6), we know that the sum of agreements, disagreements, and ties in Y (or *bracket tie* correction) is equals to the product between n_0 and n_1 , which is the denominator of τ_{br} in Formula (1).

$$A + D + E_Y = A + D + \sum_{l=1}^c n_{0l} \times n_{1l} = n_0 \times n_1 = 220$$

If the difference between agreements and disagreements is included as a numerator, we have Cureton's τ_{rb} calculated from Formula (1) in this random sample of 31 participants with tied data.

$$\tau_{br} = \frac{A - D}{n_0 \times n_1} = \frac{38 - 132}{220} = \frac{-94}{220} = -0.427$$

Now, τ_{br} from Formula (1) is equal to Glass' r_{br} (Formula (7)). The U statistics-based formula without tie correction (Formula (10)) also yields the same result as Somer's $d_{Y|X}$ (Formula (8)), Glass' r_{br} (Formula (7)), and Cureton's τ_{br} from Formula (1), but in absolute value.

$$|r_{rb}| = 1 - \frac{2U}{n_1 n_0} = 1 - \frac{2 \times 63}{11 \times 20} = 1 - \frac{126}{220} = 0.427 = |d_{SES|DM}| \neq |\gamma| = -\frac{47}{85} \approx 0.5529$$

From these last finding, it follows that:

$$\tau_{br} = r_{br} = d_{Y|X} \neq \gamma$$

$$\frac{A - D}{n_0 n_1} = \frac{2(\bar{r}_{Y|X=1} - \bar{r}_{Y|X=0})}{n_0 + n_1} = \frac{A - D}{A + D - E_Y} \neq \gamma$$

$$\frac{38-132}{11 \times 20} = \frac{2 \times \left(\frac{273}{20} - \frac{223}{11} \right)}{11+20} = \frac{38-132}{38+132+50} = -0.427$$

$$\neq \gamma = \frac{38-132}{38+132} \approx -0.5529$$

$$|\tau_{br}| = |r_{br}| = |d_{Y|X}| = 1 - \frac{2U}{n_0 n_1} \neq |\gamma|$$

$$= \left| \frac{38-132}{11 \times 20} \right| = \left| \frac{2 \times \left(\frac{273}{20} - \frac{223}{11} \right)}{11+20} \right| = \left| \frac{38-132}{38+132+50} \right| = 1 - \frac{2 \times 63}{11 \times 20} = 0.427$$

$$\neq |\gamma| = -\frac{47}{85} \approx 0.5529$$

See these equivalences applied to the examples of 10 and 7 elements.

$$A + D + E_Y = n_0 \times n_1$$

$$21 + 2 + 1 = 4 \times 6 = 24$$

$$\tau_{br} = r_{br} = d_{Y|X} \neq \gamma$$

$$\frac{A-D}{n_0 n_1} = \frac{2(\bar{r}_{Y|X=1} - \bar{r}_{Y|X=0})}{n_0 + n_1} = \frac{A-D}{A+D-E_Y} \neq \gamma$$

$$\frac{21-2}{4 \times 6} = \frac{2 \times \left(\frac{42.5}{6} - \frac{12.5}{4} \right)}{4+6} = \frac{21-2}{21+2+1} = 0.791\bar{6} \neq \gamma = \frac{21-2}{21+2} = \frac{19}{23} \approx 0.8261$$

$$|\tau_{br}| = |r_{br}| = |d_{Y|X}| = 1 - \frac{2U}{n_0 n_1} \neq |\gamma|$$

$$= \left| \frac{21-2}{4 \times 6} \right| = \left| \frac{2 \times \left(\frac{42.5}{6} - \frac{12.5}{4} \right)}{4+6} \right| = \left| \frac{21-2}{21+2+1} \right| = 1 - \frac{2 \times 2.5}{4 \times 6} = 0.791\bar{6} \neq |\gamma| \approx 0.8261$$

$$A + D + E_Y = n_0 \times n_1$$

$$7 + 2 + 3 = 3 \times 4 = 12$$

$$\tau_{br} = r_{br} = d_{Y|X} \neq \gamma$$

$$\frac{A-D}{n_0 n_1} = \frac{2(\bar{r}_{Y|X=1} - \bar{r}_{Y|X=0})}{n_0 + n_1} = \frac{A-D}{A+D-E_Y} \neq \gamma$$

$$\frac{7-2}{3 \times 4} = \frac{2 \times \left(\frac{18.5}{4} - \frac{9.5}{3} \right)}{3+4} = \frac{7-2}{7+2+3} = 0.41\bar{6} \neq |\gamma| = \frac{7-2}{7+2} = 0.5$$

$$|\tau_{br}| = |r_{br}| = |d_{Y|X}| = 1 - \frac{2U}{n_0 n_1} \neq |\gamma|$$

$$= \left| \frac{7-2}{3 \times 4} \right| = \left| \frac{2 \times \left(\frac{18.5}{4} - \frac{9.5}{3} \right)}{3+4} \right| = \left| \frac{7-2}{7+2+3} \right| = 1 - \frac{2 \times 3.5}{3 \times 4} = 0.41\bar{6} \neq |\gamma| = 0.5$$

A question arises as to which equivalence (with gamma or with asymmetric d) is more appropriate. Equivalence with the Goodman-Kruskal gamma [14] allows a non-directional estimation, which is the objective of a correlation, whereas the Somers [17] $d_{Y|X}$ implies directionality (ranking with respect to dichotomy). In this clinical example, such directionality would go against the expected, since the sociodemographic variable (ranking variable) is considered a risk factor for diabetes mellitus (dichotomy) [19]. Consequently, under the correlation proposal, equivalence with Goodman-Kruskal's gamma is more appropriate than with Somers' $d_{Y|X}$ for estimating the rank-biserial correlation. On the other hand, the formulas with the correction for ties, when present, yield higher values than without the correction.

$$\sum_{l=1}^c n_{0l}n_{1l} \neq 0 \Rightarrow \frac{A-D}{n_0 \times n_1 - \sum_{l=1}^c n_{0l}n_{1l}} > \frac{A-D}{n_0 \times n_1}$$

In this connection, Metsämuuronen [20] advises the use of Somers' d to measure the association between dichotomous items and an ordinal scale test. In this example, the items determine the total score of the test. Consequently, the ordinal variable conditional on the dichotomous variable ($Y|X$) makes sense. However, the author notes that the coefficient $d_{Y|X}$ underestimates the association between the variables, as is the case when using the formulas to calculate the rank-biserial correlation without tie correction when ties are present. In clinical studies on breast cancer [21] and COVID cases [22], the direction of the relationship is reversed ($X|Y$) and the value of the correlation is usually lower compared to the $Y|X$ direction and the bidirectional XY association. In this case, the non-directionality is clearly better.

Some websites [23] advise calculating the rank-biserial correlation using Spearman's rank correlation, which is an option offered by SPSS and other statistical software, and some studies follow this suggestion [22]. However, it is incorrect. It gives a very different and lower result due to the large number of ties in X . Cureton's tau-type formula for the rank-biserial correlation [1] [3] is equivalent to Kruskal-Wallis' gamma [14] and Glass' rho-type formula to Somers' $d_{Y|X}$ [17]. Both of them are options also present in SPSS and their use for calculating the rank-biserial correlation is more appropriate than Spearman's correlation. For example, Spearman's rho is 0.683 in Cureton's 10-element sample [1] [3], 0.378 in the 7-element sample, and -0.371 in the 31-participant sample, which are values less than Cureton's rank-biserial correlation (0.826, 0.556, and -0.553 , respectively).

9. Conclusions

Let Y be a ranking from 1 to n or a variable with k ordered categories and X a dichotomy (0 and 1, assigning 0 to the group with the lowest Y -rank sum and 1 to the group with the highest Y -rank sum). When there are no ties for Y in the two groups of X , Cureton's tau-type, Glass' rho-type and Willson's U -statistic-based formulas to calculate the rank-biserial correlation, as well as Kruskal-

Wallis' gamma and Somers' delta give the same result. These calculations are very easy to perform from a $2 \times k$ contingency table. When there are ties in the variable with k ordered categories that appear in both groups of X or *bracket ties*, not all coefficients coincide. Cureton [3] in his 1956 paper uses a correction based on the *bracket ties* when calculating the rank biserial correlation and, in 1968, Cureton [1] provided a new correction so that his tau-type formula and Glass's rho-type formula give the same result taking into account that there are ties that cause an underestimation of the relationship. In the present work, this new Cureton correction has been called correction based on a *two-sided or riding tie* to be able to differentiate it. Willson [2] applied the *two-sided tie* correction to the U-statistic-based formula with the same goal, namely to give the same result as the tau-type formula with a correction based on *bracket ties*. The only proof of the correction based on a *two-sided tie* is a 10-data example given by Cureton [1], but there is no algebraic derivation. In the present work, it is verified that the correction does not work well with an example very similar to the one given by Cureton and in another example with a sample of 31 participants. Starting from an equality given by Glass [4] in his proof that the tau-type and rho-type coefficients coincide when there are no *bracket ties*, a new correction based on the *bracket ties* is derived. It is verified with the example of Cureton [1] and the two new examples. This correction consists of dividing by two the correction given by Cureton for the tau-type formula. Thus, it is much easier than calculating the *two-sided tie* correction and is the only change required to use the formulas developed by Cureton [1] and Willson [2], achieving the initial goal that the three coefficients coincide with each other and are equivalent to Kruskal-Wallis' gamma. Hence, the asymptotic standard error based on the Mann-Whitney U test derived by Willson [2] is valid to test the significance and estimate the confidence interval with the new proposed correction. Naturally, one could change the objective and invert it, namely that all three formulas coincide with Somers' $d_{Y|X}$. This is accomplished when using the formulas without correction for ties with data that has *bracket ties*. From the point of view of a correlation or non-directional relationship, the original objective of equivalence with Kruskal-Wallis' gamma is the most correct, especially in cases where it does not make sense for the ordinal variable to be conditioned by the dichotomy, as in the clinical example shown. However, when it makes sense, equivalence with Somers' $d_{Y|X}$ might be preferred, for example, when estimating the correlation between the dichotomous items that make up a scale and the total scale score.

The rank-biserial correlation does not necessarily require the dichotomization of a ranking and can be used to estimate the linear association between a dichotomous qualitative variable and an ordinal one [3] [4], as well as to estimate the effect size with the Mann-Whitney U test [5] and the fit of a logistic regression model. Another option for these same objectives is to use the polyserial correlation proposed by Metsämuuronen [24]. However, the rank-biserial correlation formulas are much easier to calculate and yield very similar estimates [11]. For example, they can be computed with Excel or MATLAB. It is also possible to

take advantage of the equivalences with Kruskal-Wallis' gamma (with correction for ties when present) and Somers' d_{MX} (without correction for ties when present) to obtain the rank-biserial correlation, since both statistics are computed by statistical packages, such as R, Real Statistics Resource Pack, SPSS, SAS, STATA, STATISTICA, Minitab, Eviews, JASP, and others. It is important to note that using Spearman's rank correlation does not give the same result as Cureton's and Glass' formulas, but rather underestimates the linear association.

For further research, using a simulation study, it is suggested to compare the coverage, power and efficiency of confidence intervals for the rank-biserial correlation (with the new proposed correction in case of ties) based on the approximation from the Mann-Whitney U test proposed by Willson [2] versus bootstrap confidence intervals. In turn, the confidence intervals developed for Kruskal-Wallis' gamma and the Somers' d_{MX} consider can be considered [25]. The results of this type of study can provide guidance on how to implement the calculation of the rank-biserial correlation in statistical software.

Acknowledgements

The author would like to thank the referee for their helpful comments.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Cureton, E.E. (1968) Rank-Biserial Correlation when Ties Are Present. *Educational and Psychological Measurement*, **28**, 77-79. <https://doi.org/10.1177/001316446802800107>
- [2] Willson, V.L. (1976) Critical Values of the Rank-Biserial Correlation Coefficient. *Educational and Psychological Measurement*, **36**, 297-300. <https://doi.org/10.1177/001316447603600207>
- [3] Cureton, E.E. (1956) The Rank-Biserial Correlation. *Psychometrika*, **21**, 287-290. <https://doi.org/10.1007/BF02289138>
- [4] Glass, G.V. (1966) Note on Rank-Biserial Correlation. *Educational and Psychological Measurement*, **26**, 623-631. <https://doi.org/10.1177/001316446602600307>
- [5] Mann, H.B. and Whitney, D.R. (1947) On a Test of Whether One or Two Random Variables Is Stochastically Larger than the Other. *Annals of Mathematical Statistics*, **18**, 50-60. <https://doi.org/10.1214/aoms/1177730491>
- [6] Brogden, H.E. (1949) A New Coefficient: Application to Biserial Correlation and to Estimation of Selective Efficiency. *Psychometrika*, **14**, 169-182. <https://doi.org/10.1007/BF02289151>
- [7] Stanley, J.C. (1968) An Important Similarity between Biserial R and the Brogden-Cureton-Glass Biserial R for Ranks. *Educational and Psychological Measurement*, **28**, 249-253. <https://doi.org/10.1177/001316446802800204>
- [8] Khamis, H. (2008) Measures of Association: How to Choose? *Journal of Diagnostic Medical Sonography*, **24**, 155-162. <https://doi.org/10.1177/8756479308317006>

- [9] Berry, K.J., Johnston, J.E. and Mielke Jr., P.W. (2018) The Measurement of Association. A Permutation Statistical Approach. Springer, Cham.
<https://doi.org/10.1007/978-3-319-98926-6>
- [10] Kraemer, H.C. (2006) Biserial Correlation. In: Kotz, S., Read, C.B., Balakrishnan, N., Vidakovic, B. and Johnson, N.L., Eds., *Encyclopedia of Statistical Sciences*, John Wiley & Sons, Inc., Hoboken, 276-279.
<https://doi.org/10.1002/0471667196.ess0153.pub2>
- [11] Metsämuuronen, J. (2022) Rank-Polyserial Correlation: A Quest for a “Missing” Coefficient of Correlation. *Frontiers in Applied Mathematics and Statistics*, **8**, 914-932.
<https://doi.org/10.3389/fams.2022.914932>
- [12] Beck, C.T. (1986) Use of Nonparametric Correlation Analysis in Graduate Students’ Research Projects. *Journal of Nursing Education*, **25**, 41-42.
<https://doi.org/10.3928/0148-4834-19860101-14>
- [13] Kerby, D.S. (2014) The Simple Difference Formula: An Approach to Teaching Nonparametric Correlation. *Comprehensive Psychology*, **3**, Article No. 1.
<https://doi.org/10.2466/11.IT.3.1>
- [14] Goodman, L.A. and Kruskal, W.H. (1954) Measures of Association for Cross Classifications. *Journal of the American Statistical Association*, **49**, 732-764.
<https://doi.org/10.1080/01621459.1954.10501231>
- [15] Newson, R. (2008) Identity of Somers’ D and the Rank Biserial Correlation Coefficient. <https://www.rogernewsonresources.org.uk/miscdocs/ranksum1.pdf>
- [16] Newson, R. (2022) Interpretation of Somers’ D under Four Simple Models.
<https://www.rogernewsonresources.org.uk/miscdocs/intsomd1.pdf>
- [17] Somers, R.H. (1962) A New Asymmetric Measure of Association for Ordinal Variables. *American Sociological Review*, **27**, 799-811. <https://doi.org/10.2307/2090408>
- [18] Cohen, J. (1988) *Statistical Power Analysis for the Behavioral Sciences*. 2nd Edition, Lawrence Erlbaum and Associates, Hillsdale.
- [19] Volaco, A., Cavalcanti, A.M., Filho, R.P. and Précoma, D.B. (2018) Socioeconomic Status: The Missing Link between Obesity and Diabetes Mellitus? *Current Diabetes Reviews*, **14**, 321-326. <https://doi.org/10.2174/1573399813666170621123227>
- [20] Metsämuuronen, J. (2020) Somers’ D as an Alternative for the Item-Test and Item-Rest Correlation Coefficients in the Educational Measurement Settings. *International Journal of Educational Methodology*, **6**, 207-221.
<https://doi.org/10.12973/ijem.6.1.207>
- [21] Porterhouse, M.D., Paul, S., Lieberenz, J.L., Stempel, L.R., Levy, M.A. and Alvarado, R. (2022) Black Women Are Less Likely to Be Classified as High-Risk for Breast Cancer Using the Tyrer-Cuzick 8 Model. *Annals of Surgical Oncology*, **29**, 6419-6425.
<https://doi.org/10.1245/s10434-022-12140-9>
- [22] Khan, M.I., Mehmood, M., Husain, S.O., Waqar, S., Asim, M. and Rehman, N. (2021) Correlation of Severity of Disease and Changes in Basic Hematological Parameters in Patients of COVID-19. *Journal of Medical Sciences*, **29**, 69-73.
<https://doi.org/10.52764/jms.21.29.3.1>
- [23] Heidel, E. (2022) Rank Biserial Correlation between Dichotomous and Ordinal Variables. Statistics. <https://www.scalestatistics.com/rank-biserial.html>
- [24] Metsämuuronen, J. (2019) Rank Polyserial Correlation.
<http://dx.doi.org/10.13140/RG.2.2.31217.53608>
- [25] IBM (2019) IBM SPSS Statistics 24 Algorithms. SPSS Inc., Chicago.