

Intrinsically Ties Adjusted Partial Tau (C-Tap) Correlation Coefficient

Oyeka Ikewelugo Cyprian Anaene, Osuji George Amaeze, Obiora-Ilouno Happiness Onyebuchi*

Department of Statistics, Physical Science Faculty, Nnamdi Azikiwe University, Awka, Nigeria

Email address:

obiorailounoho@yahoo.com (Obiora-Ilouno H. O.)

*Corresponding author

To cite this article:

Oyeka Ikewelugo Cyprian Anaene, Osuji George Amaeze, Obiora-Ilouno Happiness Onyebuchi. Intrinsically Ties Adjusted Partial Tau (C-Tap) Correlation Coefficient. *American Journal of Theoretical and Applied Statistics*. Vol. 5, No. 5, 2016, pp. 270-279.

doi: 10.11648/j.ajtas.20160505.14

Received: December 21, 2015; **Accepted:** May 24, 2016; **Published:** August 10, 2016

Abstract: This paper present a non-parametric statistical method for the estimation of partial correlation coefficient intrinsically adjusted for tied observations in the data. The method based on a modification of the method of estimating Tau correlation coefficient may be used when the population of interest are measurements on as low as the ordinal scale that are not necessary continuous or even numeric. The estimated partial correlation coefficient is a weighted average of the estimates obtained when each of the observations whose assigned ranks are arranged in their natural order as well as the observations whose assigned ranks are tagged along, with the weights being functions of the number of tied observations in each population. It is shown that failure to adjust for ties tends to lead to an underestimation of the true partial correlation coefficient, an effect that increases with the number of ties in the data. The proposed method is illustrated with some data and shown to compare favorably with the Kendall approach.

Keywords: Intrinsically, Ties, Adjusted Partial Tau (C-Tap), Correlation, Coefficient, Estimation

1. Introduction

[4] proposed a non-parametric method for estimating the simple correlation coefficient as well as any desired partial correlation coefficient between any two samples drawn from any two populations while holding at constant level values of observations drawn from the third population.

In Kendall's approach, the populations of interest may be measurements on as low as the ordinal scale and need not be continuous or normally distributed.

Using the foundation of [4], [1] proposed a more formulated non parametric statistical method for the estimation of a partial correlation coefficient between two variables X and Y say, when the third variable Z say is held constant and again like Kendall's approach, these variables may not necessarily be continuous or normally distributed.

In this paper, we propose another non-parametric method for the estimation of partial correlation coefficient that is intrinsically adjusted for tied observations in the data.

Apart from the fact that data may not be necessarily

continuous or normally distributed, it may also not be numeric. The proposed method is more generalized and covers cases in which there are tied observations in any of the three populations which is intrinsically adjusted for those tied observation and when there are equal and unequal observation on the sampled populations which [4] and [1] did not put into consideration.

Now,

$$S_{\max} = \binom{n}{2} = \frac{n(n-1)}{2} \quad (1)$$

This maximum possible total number of agreement between the ranks assigned to observations from population X and Y provided that ranking of observations from Y and X are in their natural order [4].

$$r_k = \frac{S_a}{S_{\max}} \quad (2)$$

This is the Kendall's tau correlation coefficient rationale

and is the basis of estimation where S_a is the total sum of 1s or (+) and -1s or (-) obtained by numbers of each pair of ranks assigned to Y say, when the observation are arranged in natural ordered ranks of observations from X.

Using equation 2 for any two equal sample observation, we estimate a partial correlation coefficient between two populations X and Y say, when the values of a third population Z say are constant.

This is expressed for obtaining partial correlation coefficient based on the simple tau correlation coefficient [4], [5], [7].

$$r_{xy.z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}} \tag{3}$$

Where r_{xy} , r_{xz} and r_{yz} are respectively the tau correlation coefficient between observations from population X and Y, X and Z, Y and Z. [4], [5] methods of obtaining $r_{xy.z}$ exist in a very tedious way in practice. Again Kendall basic formulae do not have provision for presence of ties in the data.

Here, we proposed to develop using equations 2 and 3 a more formulated non parametric statistical method for the estimation of partial correlation coefficient between two variables X and Y say, when a third variable Z say is held constant, and these variables are not necessarily continuous or normally distributed, but measured on at least the ordinal scale.

The proposed method is more generalized and covers cases in which there are no tied observations in any of the three populations, and when there are equal and unequal tied observations on the on the sampled population.

The method is referred to as C-TAP for ties adjusted partial tau correlation coefficients which differentiate the proposed method from the usual Kendall tau partial correlation coefficient.

2. The Proposed Method

Consider the variables x_i, y_i and z_i for the i^{th} observation in a random sample of size n sampled from populations X, Y and Z respectively. In this method, population X, Y and Z may be measurements on as low as the ordinal scales that are not necessarily continuous or even numeric. We define r_{ix} as the rank assigned to x_i , r_{iy} the rank assigned to y_i and r_{iz} the rank assigned to z_i for $i = 1, 2, 3, \dots, n$ as usual, tied observations in each sample are assigned to their mean ranks. Then

$$U_{jk;x,z} = \begin{cases} 1, & \text{if } r_{jx} < r_{kx} \\ 0, & \text{if } r_{jx} = r_{kx} \\ -1, & \text{if } r_{jx} > r_{kx} \end{cases} \tag{4}$$

for $j = 1, 2, \dots, n-1; k = 2, 3, \dots, n; j < k$. See [1]

That is provided that the rank assigned to the k^{th}

observation from the population X comes after, that is succeeds the rank assigned to the j^{th} observation from the same population when these observations are arranged in accordance with the natural ordering or ranking of the corresponding sister observations from population Z ($j < k$).

Let

$$\begin{aligned} z^+ \pi_x &= P(u_{jk;x,z} = 1); \\ z^0 \pi_x &= P(u_{jk;x,z} = 0); \\ z^- \pi_x &= P(u_{jk;x,z} = -1) \end{aligned} \tag{5}$$

where

$$z^+ \pi_x + z^0 \pi_x + z^- \pi_x = 1 \tag{6}$$

and $z^0 \pi_x = 0$, if there are no ties in X.

Now let

$$zS_x = \sum_{j=1}^{n-1} \sum_{k=2}^n u_{jk;x,z} \tag{7}$$

$j < k$

Now from Equations (4) and (5), we have that

$$\begin{aligned} E(u_{jk;x,z}) &= z^+ \pi_x - z^- \pi_x; \text{Var}(u_{jk;x,z}) \\ &= z^+ \pi_x + z^- \pi_x - \left(z^+ \pi_x - z^- \pi_x \right)^2 \end{aligned} \tag{8}$$

Similarly;

$$E(zS_x) = \sum_{j=1}^{n-1} \sum_{k=2}^n E(u_{jk;x,z}) \tag{9}$$

$j < k$

That is

$$E(zS_x) = \frac{1}{2} (n^2 - n) \left(z^+ \pi_x - z^- \pi_x \right) = \frac{1}{2} n(n-1) \left(z^+ \pi_x - z^- \pi_x \right) \tag{9a}$$

And

$$\text{Var}(zS_x) = \frac{1}{2} n(n-1) \left(z^+ \pi_x + z^- \pi_x - \left(z^+ \pi_x - z^- \pi_x \right)^2 \right) \tag{9b}$$

Note that $z^+ \pi_x$, $z^0 \pi_x$, and $z^- \pi_x$, are respectively the probabilities that the rank assigned to the j^{th} observation from population X is less than, equal to or greater than the rank assigned to the k^{th} observation from the same population if the rank assigned to this k^{th} observation succeeds the rank assigned to the j^{th} observation from X when these observations are arranged in accordance with natural ordering of the ranks assigned to the observations from

population Z . These probabilities are estimated as

$$z\pi_x^+ = \frac{zf^+x}{\binom{n}{2}} = \frac{2zf^+x}{n(n-1)}; z\pi_x^0 = \frac{2zf^0x}{n(n-1)}; z\pi_x^- = \frac{2zf^-x}{n(n-1)} \quad (10)$$

where zf^+x , zf^0x and zf^-x are respectively the number of 1s, 0s, and -1s in the frequency distribution of these numbers in $ujk; x.z; j = 1, 2, \dots, n-1, k = 2, 3, \dots, n, j < k$.

Hence the sample estimate of the total number of times the ranking of observations from population X are in their natural order and consistent with the natural order and consistent with the natural ordering of the ranks of observations from population Z less the number of times they are out of order is obtained from Equation (9) and (10) as

$$zSx = \frac{1}{2}n(n-1)\left(z\pi_x^+ - z\pi_x^-\right) = zf^+x - zf^-x \quad (11)$$

As noted above, if these rankings are in their natural order then the maximum possible total number of arrangements or scores is $S_{max} = \binom{n}{2}$ (See Eqn (1)). Hence the Kendall tau correlation coefficient between observations from population X and observations from population Z uncorrected for ties in X may be estimated using Equations (i) and (ii) in Equations 2 as

$$zrx = \frac{zSx}{S_{max}} = \frac{zSx}{n(n-1)/2} \text{ or } zrx = \frac{zf^+x - zf^-x}{n(n-1)} \quad (12)$$

Note that if there are tied observations in X , then the estimated correlation coefficient zrx of Eqn 12 would not be unbiased estimate of the true tau correlation coefficient between X and Z . This is because even though the numerator zSx has by specifications been adjusted for possible ties in X , its denominator S_{max} has not been so

$$z\pi_y^+ = P(ujk; y.z = 1); z\pi_y^0 = P(ujk; y.z = 0); z\pi_y^- = P(ujk; y.z = -1) \quad (16)$$

$$\text{where } z\pi_y^+ + z\pi_y^0 + z\pi_y^- = 1 \quad (17)$$

and $z\pi_y^+ = 0$ if there are no ties in Y .

Also let

$$zS_y = \sum_{j=1}^{n-1} \sum_{\substack{k=2 \\ j < k}}^n ujk; y.z \quad (18)$$

Now from Eqns 15 and 16, we have that

adjusted. Therefore S_{max} needs to be adjusted. To do this, we subtract f^0x , the number of tied observations in X from S_{max} to obtain

$$S_{max.c} = S_{max} - f^0x = \binom{n}{2} - f^0x \quad (13)$$

Hence an estimate of a tau correlation coefficient between X and Z adjusted for ties in X is

$$zrx.c = \frac{zSx}{S_{max.c}} = \frac{zSx}{S_{max} - f^0x} = \frac{zSx / \binom{n}{2}}{1 - f^0x / \binom{n}{2}} \text{ or}$$

$$zrx.c = \frac{zrx}{1 - z\pi_x^0} = \frac{z\pi_x^+ - z\pi_x^-}{1 - z\pi_x^0} \quad (14)$$

The unadjusted and adjusted tau correlation coefficients between observations from populations Y and Z are similarly estimated. Thus having arranged the ranks assigned to observations from Z in their natural order and tagged along the ranks assigned to the corresponding observations from Y we let

$$ujk; y.z = \begin{cases} 1, & \text{if } r_{jy} < r_{ky} \\ 0, & \text{if } r_{jy} = r_{ky} \\ -1, & \text{if } r_{jy} > r_{ky} \end{cases} \quad (15)$$

For $j = 1, 2, \dots, n-1; k = 2, 3, \dots, n, j < k$, that is, provided that the rank assigned to the k^{th} observation from population Y comes after, that it succeeds the rank assigned to the j^{th} observation from the same population when these observations are arranged in accordance with the natural ordering or ranking of the corresponding sister observation from population $Z (j < k)$.

Let

$$E(ujk; y.z) = z\pi_y^+ - z\pi_y^- : Var(ujk; y.z) = z\pi_y^+ + z\pi_y^- - (z\pi_y^+ - z\pi_y^-)^2 \tag{19}$$

Similarly, from Eqns 18 and 19

$$E(zS_y) = \frac{1}{2}n(n-1)\left(z\pi_y^+ - z\pi_y^-\right) \tag{20}$$

And

$$Var(zS_y) = \frac{n(n-1)}{2}\left(z\pi_y^+ + z\pi_y^- - \left(z\pi_y^+ - z\pi_y^-\right)^2\right) \tag{21}$$

Note that $z\pi_y^+, z\pi_y^0, z\pi_y^-$, and are respectively the probabilities that the rank assigned to the j^{th} observation from population Y is less than, equal to or greater than the rank assigned to k^{th} observation from the same population if the rank assigned to the k^{th} observation succeeds the rank assigned to the j^{th} observation when the ranks assigned to these observations are arranged in accordance with the natural ordering of the ranks assigned to the observations from population Z . These probabilities are estimated as

$$z\pi_y^+ = \frac{2zf^+y}{n(n-1)}; z\pi_y^0 = \frac{2zf^0y}{n(n-1)}; z\pi_y^- = \frac{2zf^-y}{n(n-1)} \tag{22}$$

Where zf^+y, zf^0y and zf^-y are respectively the number of 1s, 0s, and -1s in the frequency distribution of these numbers in $ujk; y.z; j = 1, 2 \dots n-1; k = 2, 3 \dots n, j < k$.

Hence the sample estimate of the total number of times the rankings of observations from population Y are in their natural order and consistent with the natural ordering of the ranks of observations from population Z less the number of times they are out of order is obtained from Equations (20) and (22) as

$$zS_y = \frac{1}{2}n(n-1)\left(z\pi_y^+ - z\pi_y^-\right) = zf^+y - zf^-y \tag{23}$$

Hence as before, the tau correlation coefficient between Y and Z unadjusted for ties in Y is estimated as

$$zry = \frac{zS_y}{S_{\max}} = \frac{\frac{1}{2}n(n-1)\left(z\pi_y^+ - z\pi_y^-\right)}{\binom{n}{2}} \text{ or}$$

$$\pi_{xy}^+ = P(ujk; xy.z = 1); \pi_{xy}^0 = P(ujk; xy.z = 0); \pi_{xy}^- = P(ujk; xy.z = -1) \tag{29}$$

$$\text{where } \pi_{xy}^+ + \pi_{xy}^0 + \pi_{xy}^- = 1 \tag{30}$$

and $\pi_{xy}^0 = 0$ if there are no ties between X and Y

Define

$$zry = z\pi_y^+ - z\pi_y^- = \frac{2(zf^+y - zf^-y)}{n(n-1)} \tag{24}$$

The corresponding estimate for variance is from Eqn (21)

$$Var(zry) = \frac{2\left(z\pi_y^+ + z\pi_y^- - \left(z\pi_y^+ - z\pi_y^-\right)^2\right)}{n(n-1)} \tag{25}$$

Now the tau correlation coefficient between Y and Z adjusted for ties in Y is estimated as

$$zry.c = \frac{zS_y}{S_{\max}} = \frac{zS_y}{S_{\max} - f^0y} = \frac{zS_y / \binom{n}{2}}{1 - f^0y / \binom{n}{2}} \text{ or}$$

$$zry.c = \frac{zry}{1 - z\pi_y^0} = \frac{z\pi_y^+ - z\pi_y^-}{1 - z\pi_y^0} \tag{26}$$

whose estimated variance using Eqn (25) is

$$Var(zry.c) = \frac{2\left(z\pi_y^+ + z\pi_y^- - \left(z\pi_y^+ - z\pi_y^-\right)^2\right)}{n(n-1)\left(1 - z\pi_y^0\right)^2} \tag{27}$$

To estimate the tau correlation coefficient between observations from population X and observations from population Y , we let

$$ujk; xy.z = \begin{cases} 1, & \text{if } rjx < rky \\ 0, & \text{if } rjx = rky \\ -1, & \text{if } rjx > rky \end{cases} \tag{28}$$

for $j = 1, 2 \dots n-1; k = 2, 3 \dots n, j < k$. That is provided that the rank assigned to the k^{th} observation in the sample drawn from population Y comes after, that is, succeeds the rank assigned to the j^{th} observation in the sample drawn from population X when the ranks assigned to the observations have been arranged according to the natural ordering of the ranks assigned to their sister observations in the sample drawn from population $Z, j = 1, 2 \dots n-1; k = 2, 3 \dots n, j < k$. Also let

$$S_{xy} = \sum_{j=1}^{n-1} \sum_{k=2}^n ujk; xy.z \tag{31}$$

$j < k$

from Eqns (28) and (29), we have that

$$E(ujk; xy.z) = \pi^+ xy - \pi^- xy; \text{Var}(ujk; xy.z) = \pi^+ xy + \pi^- xy - \left(\pi^+ xy - \pi^- xy\right)^2 \tag{32}$$

Similarly from Equations (31) and (32) we have

$$E(Sxy) = \frac{1}{2} n(n-1) \left(\pi^+ xy - \pi^- xy\right) \tag{33}$$

$$\text{and } \text{Var}(Sxy) = \frac{n(n-1)}{2} \left(\pi^+ xy + \pi^- xy - \left(\pi^+ xy - \pi^- xy\right)^2\right) \tag{34}$$

Note that $\pi^+ xy, \pi^0 xy$ and $\pi^- xy$ are respectively the probabilities that the rank assigned to the j^{th} observation from population X is less than, equal to or greater than the rank assigned to the k^{th} observation from population Y , when the observations have been arranged so that the rank assigned to them correspond with the natural order of the ranks of their sister observations from population Z with the rank assigned to the k^{th} observation in Y succeeding the ranks assigned to the j^{th} observations in X and are estimated as

$$\pi^+ xy = \frac{2f^+ xy}{n(n-1)}; \pi^0 xy = \frac{2f^0 xy}{n(n-1)}; \pi^- xy = \frac{2f^- xy}{n(n-1)} \tag{35}$$

where $f^+ xy, f^0 xy$ and $f^- xy$ and are respectively the number of 1s, 0s, and -1s in the frequency distribution of these numbers in $ujk; xy.z, j = 1, 2, \dots, n-1; k = 2, 3, \dots, n; j < k$.

Hence the sample estimate of the total number of times the rankings of observations from population X and Y are in their natural order and consistent with the natural ordering with their sister observations from population Z less the number of times they are out of order is estimated from Equations (33) and (35) as

$$Sxy = \frac{1}{2} n(n-1) \left(\pi^+ xy - \pi^- xy\right) = f^+ xy - f^- xy \tag{36}$$

Following a similar argument as above the tau correlation coefficient between X and Y unadjusted for ties between these populations is estimated as

$$r_{xy} = \frac{Sxy}{S_{\max}} = \frac{\frac{1}{2} n(n-1) \left(\pi^+ xy - \pi^- xy\right)}{\binom{n}{2}} \text{ or } r_{xy} = \frac{\pi^+ xy - \pi^- xy}{2} = \frac{2(f^+ xy - f^- xy)}{n(n-1)} \tag{37}$$

whose variance is estimated as

$$\text{Var}(r_{xy}) = \frac{2 \left(\pi^+ xy + \pi^- xy - \left(\pi^+ xy - \pi^- xy\right)^2\right)}{n(n-1)} \tag{38}$$

The corresponding ties-adjusted tau correlation coefficient for ties between X and Y is estimated as

$$r_{xy.c} = \frac{Sxy}{S_{\max.c}} = \frac{Sxy}{S_{\max} - f^0 xy} = \frac{Sxy / \binom{n}{2}}{1 - f^0 xy / \binom{n}{2}}$$

That is

$$r_{xy.c} = \frac{r_{xy}}{1 - \pi^0 xy} = \frac{\pi^+ xy - \pi^- xy}{1 - \pi^0 xy} \tag{39}$$

The corresponding variance is estimated from Equation 38 as

$$\text{Var}(r_{xy.c}) = \frac{2 \left(\pi^+ xy + \pi^- xy - \left(\pi^+ xy - \pi^- xy\right)^2\right)}{n(n-1) \left(1 - \pi^0 xy\right)^2} \tag{40}$$

As noted above, in the presence of ties in the data, the estimated tau correlation coefficient is not independent of which of the two populations being correlated has its assigned ranks arranged in their natural order and which has its assigned ranks tagged along. To adjust for this effect we would need to use each of these two sets of ranks to alternatively play each of the two roles. Thus to estimate r_{xz} , the correlation coefficient between X and Z when X has its assigned ranks arranged in their natural order, and Z has its corresponding assigned ranks tagged along, we let

$$ujk; z.x = \begin{cases} 1, & \text{if } rjz < rkz \\ 0, & \text{if } rjz = rkz \\ -1, & \text{if } rjz > rkz \end{cases} \tag{41}$$

for $j = 1, 2, \dots, n-1; k = 2, 3, \dots, n; j < k$

Also let

$$x^+ \pi z = P(ujk; z.x = 1); x^0 \pi z = P(ujk; z.x = 0); x^- \pi z = P(ujk; z.x = -1) \tag{42}$$

$$\text{where } x^+ \pi z + x^0 \pi z + x^- \pi z = 1 \tag{43}$$

and $x\pi_z = 0$; if there are no ties in Z

$$\text{Also let } xS_z = \sum_{j=1}^{n-1} \sum_{k=2}^n u_{jk}; z.x \quad (44)$$

$$j < k$$

Now

$$E(u_{jk}; z.x) = x\pi_z^+ - x\pi_z^-; \text{Var}(u_{jk}; z.x) = x\pi_z^+ + x\pi_z^- - (x\pi_z^+ - x\pi_z^-)^2 \quad (45)$$

$$\text{And } E(xS_z) = \frac{1}{2}n(n-1)(x\pi_z^+ - x\pi_z^-) \quad (46)$$

$$\text{Var}(xS_z) = \frac{1}{2}n(n-1)\left(x\pi_z^+ + x\pi_z^- - (x\pi_z^+ - x\pi_z^-)^2\right) \quad (47)$$

Note that $x\pi_z^+$, $x\pi_z^0$ and $x\pi_z^-$ are respectively the probabilities that the rank assigned to the j^{th} observation from population Z is less than, equal to or greater than the rank assigned to the K^{th} observation from the same population if the rank assigned to the K^{th} observation succeeds the rank assigned to the j^{th} observation when the ranks assigned to these observations are arranged in accordance with the natural ordering of the ranks assigned to the corresponding sister observations from population X and are estimated as

$$x\pi_z^+ = \frac{2xf^+z}{n(n-1)}; x\pi_z^0 = \frac{2xf^0z}{n(n-1)}; x\pi_z^- = \frac{2xf^-z}{n(n-1)} \quad (48)$$

where xf^+z , xf^0z and xf^-z are respectively the number of 1s, 0s, and -1s in the frequency distribution of the $\frac{1}{2}n(n-1)$ values of these numbers in $u_{jk}; x.z = 1, 2, \dots, n-1, k = 2, 3, \dots, n, j < k$.

Hence the sample estimate of the total number of times the ranking of observations from population Z are in their natural order and consistent with the natural ordering of the ranks of observations from population X less the number of times they are out of order is obtained from equation 46 and 48 as

$$xS_z = \frac{1}{2}n(n-1)(x\pi_z^+ - x\pi_z^-) = xf^+z - xf^-z \quad (49)$$

Hence as before, the tau correlation coefficient between X and Z unadjusted for ties in Z is estimated as

$$xr_z = \frac{xS_z}{S_{\max}} = \frac{\frac{1}{2}n(n-1)(x\pi_z^+ - x\pi_z^-)}{\binom{n}{2}} \text{ or}$$

$$xr_z = x\pi_z^+ - x\pi_z^- = \frac{2(xf^+z - xf^-z)}{n(n-1)} \quad (50)$$

with estimated variance

$$\text{Var}(xr_z) = \frac{2\left(x\pi_z^+ + x\pi_z^- - (x\pi_z^+ - x\pi_z^-)^2\right)}{n(n-1)} \quad (51)$$

Now the tau correlation coefficient between X and Z adjusted for ties in Z is estimated as

$$xr_{z.c} = \frac{xS_z}{S_{\max}} = \frac{xS_z}{S_{\max} - xf^0z} = \frac{xS_z / \binom{n}{2}}{1 - xf^0z / \binom{n}{2}}$$

That is

$$xr_{z.c} = \frac{xr_z}{1 - x\pi_z^0} = \frac{x\pi_z^+ - x\pi_z^-}{1 - x\pi_z^0} \quad (52)$$

whose estimated variance using equation 51 is

$$\text{Var}(xr_{z.c}) = \frac{2\left(x\pi_z^+ + x\pi_z^- - (x\pi_z^+ - x\pi_z^-)^2\right)}{n(n-1)\left(1 - x\pi_z^0\right)^2} \quad (53)$$

The tau correlation coefficients between populations Y and Z unadjusted as well as adjusted for ties in Z are similarly estimated. Thus following the above procedures we have that

$$yr_z = \frac{yS_z}{S_{\max}} = \frac{yS_z}{\binom{n}{2}} \text{ or } yr_z = y\pi_z^+ - y\pi_z^- = \frac{2(yf^+z - yf^-z)}{n(n-1)} \quad (54)$$

And

$$yr_{z.c} = \frac{yS_z}{S_{\max.c}} = \frac{yS_z}{\binom{n}{2}} - yf^0z = \frac{yS_z / \binom{n}{2}}{1 - yf^0z / \binom{n}{2}}$$

That is

$$ryz.c = \frac{y\pi z}{1-y\pi z} = \frac{y\pi z - y\pi z}{1-y\pi z} \quad (55)$$

The ties adjusted tau correlation coefficient between X and Z is estimated as a weighted average of $rxz.c$ and $zrx.c$, where the weights are functions of the number of tied observations in the two sampled populations and is estimated as

$$rxz.c = \frac{\left(1-x\pi z\right)rxz.c + \left(1-z\pi x\right)zrx.c}{2-x\pi z-z\pi x} \text{ or}$$

$$rxz.c = \frac{rxz + zrx}{2-x\pi z-z\pi x} = \frac{\left(x\pi z-x\pi z\right) + \left(z\pi x-z\pi x\right)}{2-x\pi z-z\pi x} \quad (56)$$

Similarly, the ties adjusted tau correlation coefficient between Y and Z is estimated as

$$ryz.c = \frac{y\pi z + z\pi y}{2-y\pi z-z\pi y} = \frac{\left(y\pi z-y\pi z\right) + \left(z\pi y-z\pi y\right)}{2-y\pi z-z\pi y} \quad (57)$$

Use of Equations 39, 56, and 57 in Eqn 3, yields an estimate of the ties adjusted or corrected partial tau correlation coefficient between observations from populations X and Y holding at a constant level of observations from population Z as

$$C-TAP = r_{xy.z;c} = \frac{r_{xy.c} - r_{xz.c} r_{yz.c}}{\sqrt{(1-r_{xz.c}^2)(1-r_{yz.c}^2)}} \quad (58)$$

Note that the unadjusted partial correlation coefficient $r_{xy.z}$ is equal to the ties adjusted partial correlation coefficient, $r_{xy.z;c}$ only if there are no ties in the data, otherwise $r_{xy.z}$ tends to provide an under estimate of the true partial correlation coefficient, a bias that increases with the number of ties in the data.

3. Illustrative Example

The following are the later grades earned by 13 candidates under three judges in a job interview. Two of the judges X and Y are males while judge Z is female.

Table 1. Grades of Candidates under three judges in a job interview.

candidate number	judge X (x_i)	rank under judge X (rix)	judge Y (y_i)	rank under judge Y (riy)	judge Z (z_i)	rank under judge Z (riz)
1	A	3	A ⁻	4	B ⁺	6.5
2	A ⁻	4.5	A	2	A ⁻	3
3	B ⁺	6.5	B ⁻	9.5	A ⁻	3
4	A ⁻	4.5	B ⁺	6	F	12.5
5	A ⁺	1.5	A	2	A ⁺	1
6	F	13	E	13	B ⁺	6.5
7	C ⁻	8.5	C ⁻	11.5	F	12.5
8	B ⁺	6.5	B	8	B ⁺	6.5
9	C ⁻	8.5	B ⁺	6	E	11
10	E	11	B ⁺	6	B ⁺	6.5
11	E	11	C ⁻	11.5	C ⁻	9.5
12	A ⁺	1.5	A	2	C ⁻	9.5
13	E	11	B ⁻	9.5	A ⁻	3

Interest here is to estimate the partial correlation coefficient between the score or assessment of the candidates by the two male judges when the female judge is controlled. The later grades awarded the candidates by each of the three judges are here ranked from the highest ‘A⁺’ to the lowest ‘F’, assigning the rank of 1 to A⁺, the rank of 2 to A and so on

until the rank of 13 is assigned to ‘F’. All tied grades or scores under each judge are assigned their mean rank.

To apply the proposed method, we here first arrange the ranks assigned by judge Z in their natural order and then tag along the ranks of the grades assigned the candidates by each of the other two judges X and Y the results are shown in table 2

Table 2. Natural Order of ranks for Z with corresponding ranks for X and Y.

Candidate Number	Natural Order for Rank of Z ($riz=1$)	Rank for X (rix)	Rank for Y (riy)
5	1	1.5	2
2	3	4.5	2
3	3	6.5	9.5
13	3	11	9.5
1	6.5	3	4
6	6.5	13	13
8	6.5	6.5	8
10	6.5	11	6
11	9.5	11	11.5
12	9.5	1.5	2
9	11	8.5	6
4	12.5	4.5	6
7	12.5	8.5	11.5

To calculate z_{ix} , the ties adjusted correlation coefficient between X and Z when the ranks of Z are arranged in their natural order and the corresponding ranks of X are tagged

along, we may first obtain the values of $ujk;x.z$ of Equation 4 preferably in a tabular form (Table 3)

Table 3. Calculation of Values of $ujk;x.z$ (Equation 4).

Candidate number	r_{jx}	Candidate Number	r_{kx}	5	2	3	13	1	6	8	10	11	12	9	4	7
				1.5	4.5	6.5	11	3	13	6.5	11	11	1.5	8.5	4.5	8.5
5	1.5				1	1	1	1	1	1	1	1	0	1	1	1
2	4.5					1	1	-1	1	1	1	1	-1	1	0	1
3	6.5						1	-1	1	0	1	1	-1	1	-1	1
13	11							-1	1	-1	0	0	-1	-1	-1	-1
1	3								1	1	1	1	-1	1	1	1
6	13									-1	-1	-1	-1	-1	-1	-1
8	6.5										1	1	-1	1	-1	1
10	11											0	-1	-1	-1	-1
11	11												-1	-1	-1	-1
12	1.5													1	1	1
9	8.5														-1	0
4	4.5															1
7	8.5															

From Table 3, we have that $z_{f^+x} = 41$, $z_{f^0x} = 7$ and $z_{f^-x} = 30$. Hence from Equation (10), we have

$$z_{\pi x}^+ = \frac{41}{78} = 0.526; \quad z_{\pi x}^0 = \frac{7}{78} = 0.090 \text{ and}$$

$$z_{\pi x}^- = \frac{30}{78} = 0.385$$

Hence from Equation (12), we have that the estimated tau correlation coefficient unadjusted for ties in X is $z_{rx} = 0.526 - 0.385 = 0.141$.

Therefore from Equation 14, we have that the estimated tau correlation coefficient between X and Z adjusted for ties in X is

$$z_{rx.c} = \frac{0.141}{1 - 0.090} = \frac{0.141}{0.91} = 0.155$$

The estimated tau correlation coefficient between Y and Z is obtained from the data of table 4

Table 4. Calculation of Values of $ujk;y.z$ (Equation 15).

Candidate number	r_{jy}	Candidate Number	r_{ky}	5	2	3	13	1	6	8	10	11	12	9	4	7
				2	2	9.5	9.5	4	13	8	6	11.5	2	6	6	11.5
5	2				0	1	1	1	1	1	1	1	0	1	1	1
2	2					1	1	1	1	1	1	1	0	1	1	1
3	9.5						0	-1	1	-1	-1	1	-1	-1	-1	1
13	9.5							-1	1	-1	-1	1	-1	-1	-1	1
1	4								1	1	1	1	-1	1	1	1
6	13									-1	-1	-1	-1	-1	-1	-1
8	8										-1	1	-1	-1	-1	-1
10	6											1	-1	0	0	1
11	11.5												-1	-1	-1	0
12	2													1	1	1
9	6														0	1
4	6															1
7	11.5															

From table 4, we have that

$$z_{f^+y} = 42; \quad z_{f^0y} = 8 \text{ and } z_{f^-y} = 28$$

$$z_{ry} = 0.538 - 0.359 = 0.179$$

Hence from Equation 22, we have

$$z_{\pi y}^+ = \frac{42}{78} = 0.538; \quad z_{\pi y}^0 = \frac{8}{78} = 0.103 \text{ and}$$

$$z_{\pi y}^- = \frac{28}{78} = 0.359$$

Hence the tau correlation coefficient between Y and Z adjusted for ties in Y is from equation 26

$$z_{ry.c} = \frac{0.179}{1 - 0.103} = \frac{0.179}{0.897} = 0.200$$

Therefore, using Equation 24, we have

The tau correlation coefficient r_{xy} between X and Y is obtained using the values of $ujk;xy.z$ of equation 28 shown in table 5.

Table 5. Calculation of Values of $ujk;xy$ of Equation 28.

Candidate number (rjx)	Candidate number (rky)	5	2	3	13	1	6	8	10	11	12	9	4	7
		2	2	9.5	9.5	4	13	8	6	11.5	2	6	6	11.5
5	1.5		1	1	1	1	1	1	1	1	1	1	1	1
2	4.5			1	1	-1	1	1	1	1	-1	1	1	1
3	6.5				1	-1	1	1	-1	1	-1	-1	-1	1
13	11					-1	1	-1	-1	1	-1	-1	-1	1
1	3						1	1	1	1	-1	1	1	1
6	13							-1	-1	-1	-1	-1	-1	-1
8	6.5								-1	1	-1	-1	-1	1
10	11									1	-1	-1	-1	1
11	11										-1	-1	-1	1
12	1.5											1	1	1
9	8.5												-1	1
4	4.5													1
7	8.5													

From table 5, we have that

$$f^+xy = 46; f^0xy = 0; f^-xy = 32$$

So that from Equation 35,

$$\hat{\pi}^+xy = \frac{46}{78} = 0.590; \hat{\pi}^0xy = 0.00; \hat{\pi}^-xy = \frac{32}{78} = 0.410$$

Therefore, from Equation 37, the tau correlation coefficient between X and Y unadjusted for any ties between

these two variables is estimated as

$\tau_{xy} = 0.590 - 0.410 = 0.180$, which is here the same as the adjusted correlation coefficient since there are no common tied observations between x .

To estimate the tau correlation coefficient τ_{xz} , between X and Z when the ranks assigned to observations from X are arranged in their natural order and the ranks assigned to the corresponding observations from Z are tagged along, we use the values of $ujk;z.x$ of equation 41 shown in table 6.

Table 6. Calculating the Values of $ujk;z.x$ of Equation 41.

Candidate number (rjx)	Candidate number (rky)	5	12	1	2	4	3	8	7	9	10	11	13	6
		1	9.5	6.5	3	12.5	3	6.5	12.5	11	6.5	9.5	3	6.5
5	1		1	1	1	1	1	1	1	1	1	1	1	1
12	9.5			-1	-1	1	-1	-1	1	1	-1	0	-1	-1
1	6.5				-1	1	-1	0	1	1	0	1	-1	0
2	3					1	0	1	1	1	1	1	0	1
4	12.5						-1	-1	0	-1	-1	-1	-1	-1
3	3							1	1	1	1	1	0	1
8	6.5								1	1	0	1	-1	0
7	12.5									-1	-1	-1	-1	-1
9	11										-1	-1	-1	-1
10	6.5											1	-1	0
11	9.5												-1	-1
13	3													-1
6	6.5													

From table 6, $xf^+z = 37$, $xf^0z = 11$, and $xf^-z = 30$

Hence from equation (48)

$$\hat{\pi}^+zx = \frac{37}{78} = 0.474, \hat{\pi}^0zx = \frac{11}{78} = 0.141, \text{ and}$$

$$\hat{\pi}^-zx = \frac{30}{78} = 0.385,$$

The estimated tau correlation coefficient between X and Z unadjusted for ties in Z (Eqn 50) is $\tau_{xz} = 0.474 - 0.385 = 0.089$.

The corresponding ties adjusted correlation coefficient (equation 52) is

$$\tau_{xz.c} = \frac{0.089}{1 - 0.141} = \frac{0.089}{0.859} = 0.104$$

The corresponding tau correlation coefficient between Y

and Z when the ranks assigned to observations from Y are arranged in their natural order and the ranks assigned to the observations from Z are tagged along namely, τ_{yz} is estimated using the values of $ujk;z.y$ from which we obtain and $yf^+z = 37$; $yf^0z = 11$ and $yf^-z = 30$

Hence $\hat{\pi}^+yz = \frac{37}{78} = 0.474,$ $\hat{\pi}^0yz = \frac{11}{78} = 0.141,$ and

$$\hat{\pi}^-yz = \frac{30}{78} = 0.385.$$

So that the tau correlation coefficient between Y and Z unadjusted for ties in Z is estimated as

$\tau_{yz.c} = \hat{\pi}^+yz - \hat{\pi}^-yz = 0.474 - 0.385 = 0.089$, while the corresponding ties adjusted correlation coefficient is

$$\tau_{yz.c} = \frac{\tau_{yz}}{1 - \hat{\pi}^0yz} = \frac{0.089}{1 - 0.141} = 0.104.$$

From these results, we have that the estimated ties adjusted tau correlation coefficient between X and Z from equation 56 is

$$r_{xz.c} = \frac{0.141 + 0.089}{2 - 0.090 - 0.141} = \frac{0.230}{1.769} = 0.130$$

Similarly, the estimated ties adjusted tau correlation coefficient between Y and Z (Eqn 57)

$$r_{yz.c} = \frac{0.179 + 0.089}{2 - 0.103 - 0.141} = \frac{0.268}{1.756} = 0.153$$

Using these values in equation 58, the ties adjusted partial correlation coefficient between X and Y, holding Z constant is estimated as

$$C-TAP = r_{xy.z;c} = \frac{0.180 - (0.130)(0.153)}{\sqrt{(1 - (0.130)^2)(1 - (0.153)^2)}} = \frac{0.160}{0.980} = 0.163$$

Notice that the simple correlation coefficient between X and Y is about 0.180 while the partial correlations coefficient between X and Y is only 0.163, indicating that the assessment by the female judge seems to reduce the strength of the association or agreement between the male judges in the assessment of the candidates. Also, if no adjustments have been made for the presence of tied observations in the data, the estimated partial correlation coefficient would have been

$$r_{xy.z} = \frac{0.180 - (0.141)(0.179)}{\sqrt{(1 - (0.141)^2)(1 - (0.179)^2)}} = \frac{0.155}{0.974} = 0.159$$

Showing that not adjusted for the presence of ties in the data, tends to lead to a probable underestimation of the true partial correlation coefficient.

4. Conclusion

This paper has presented a non-parametric method for estimating partial correlation coefficient between two variables holding a third variable constant, when there are tied observations in the data. The proposed method is more generalized and covers in which there are no tied observations in any of the three populations; and when there are equal and unequal tied observations on the sampled populations. The proposed method uses a modified approach to the estimation of tau correlation coefficient and assumes

that the populations of interest may be measurements on as low as the ordinal scale. The estimated ties adjusted partial correlation coefficient is a weighted average of the estimates obtained when the ranks assigned to the observations from each of the sampled populations are alternatively used as those that are naturally ordered as well as those that are tagged along. It is shown that when there are ties in the data, failure to adjust for these ties tend to result to an underestimate of the true partial correlation coefficient. This bias increase with the number of tied observations in the data.

The proposed method is illustrated with some data and is shown to compare favorably with the Kendall approach.

References

- [1] Ebuh G. U, Oyeka I. C. A (2012) "A Nonparametric Method for Estimating Partial Correlation Coefficient" J Biom Biostat 3: 156. doi:10.4172/2155-6180.1000156.
- [2] Fraser, D. A. S. (1957) "Nonparametric Methods in Statistics" John Wiley & Sons, Inc., New York.
- [3] Gibbon, J. D (1973), "Non parametric statistical inference" Mc Graw Hills book company, New York.
- [4] Hollander, m and Woife, D. A. (1999), "Non parametric statistical Methods (2nd edition) Wiley-Inter Science, New York.
- [5] Kendall, M. G. (1948), Rank Correlation Methods, Hafiner Publishing Company. Inc. New York.
- [6] Noether, G. E. (1967) "Elements of Nonparametric Statistics" John Wiley & Sons, Inc., New York.
- [7] Siegel Sidney (1956) Nonparametric Statistics for the Behavioral Sciences. McGraw-Hill Series in Psychology, New York.
- [8] Oyeka, C. A Osuji, G. A and Nwankwo, C. C. (2013), "Intrinsically ties adjusted Tau (C - Tat) Correlation Coefficient" American Journal of Theoretical and Applied Statistics Vol 2 pp 273-281.
- [9] Oyeka I. C. A., Ebuh G. U., Nwosu C. R., Utazi E. C., Ikpegbu P. A., Obiora-Ilouno H & Nwankwo C. C. (2009) "A Method of Analyzing Paired Data Intrinsically Adjusted for Ties". Global Journal of Mathematics and Statistics, India. Volume1, Number 1, 2009. 1-6.
- [10] Oyeka, C. A. (1996) "An introduction to applied statistical method. Nobern avocation Publication Company, Enugu-Nigeria.