Simultaneous Determination Of Adjusted Ranks Of Sample Observations And Their Sums And Products

Oyeka Cyprian Anene, Nwankwo Chike H

Abstract: This paper proposes a systematic method for the simultaneous determination of adjusted ranks of sample observations and their sums and products adjusted for possible presence of tied observations in the sampled populations for use in further analyses. When computations involving paired data sets, as in the computation of the Spearman Rank Correlation Coefficient, this procedure intrinsically obtained the sums of ranks, products of ranks and sums of squares of ranks, automatically adjusting these sums for more accurate results. The proposed method is illustrated with some data and used to estimate ties adjusted Spearman Rank Correlation Coefficient and the bias that would have arisen if there were no adjustments for ties in the sampled populations.

Keywords: Rank, Adjusted, Systematic, Ties, Bias, Intrinsically, Sums of Squares.

I. Introduction

Sometimes one may have a sample drawn from a population that may be a measurement on as low as the ordinal scales that are not necessarily continuous or even numeric. Research interest is in using ranks instead of raw scores in statistical analyses. In this case one may have to first assign ranks to each of the observations before further analyses. A problem that often arises is how to systematically assign ranks to these data without having to arrange them in any order, either from the smallest to the largest or from the largest to the smallest. It is acknowledged that several methods exist for generating rank sums, products, sums of squares and breaking of ties between sample observations in their rankings (Gibbons J. D. 1973; Hollander and Wolfe, 1999; Siegel Sidney, 1956; Oyeka et al, 2014) adopting different approaches. We, in this paper, propose to develop a statistical method along the line of Oyeka et al (2013) to help in systematically assigning ranks to a sample of observations without first arranging them in any order. We also develop a statistical method for the estimation of the sums of squares and products of ranks in the presence of tied observations in which some or all the observations are tied and therefore assigned mean ranks.

II. Proposed method

Let x_j be the jth observation in a random sample of size n drawn from population X_1 for j = 1, 2, …, n where x_j may be measurements on as low as the ordinal scale and need not be continuous or numeric. To assign ranks to these observations we may let.

- *Oyeka Cyprian Anene and Nwankwo Chike H*
  Department of Statistics, Faculty of Physical Sciences, Nnamdi Azikiwe University, Awka. Nigeria.
  E-mai: chikeezeoke@yahoo.com

$$U_{lj} = \begin{cases} 1, if\ l = j;\ or\ for\ l \neq j\ if\ x_j\ is\ higher\ (better\ greater) than\ x_l \\ \frac{1}{2}, if\ x_j\ is\ the\ same\ as\ (equal\ to)\ x_l \\ 0, if\ x_j\ is\ lower\ (worse, smaller) than\ x_l,\ for\ l, j = 1,2\ ....,n \end{cases}$$

… (1)

Let

$$\pi_j^+ = P(U_{lj} = 1);\ \pi_j^e = P\left(U_{lj} = \frac{1}{2}\right);\ \pi_j^o = P(U_{lj} = 0)\ .\ .\ .\ (2)$$

The expected value and variance of U_{lj} are respectively

$$E(U_{lj}) = \pi_j^+ + \frac{1}{2}\pi_j^e$$

$$Var(U_{lj}) = \pi_j^+ + \frac{1}{4}\pi_j^e - \left(\pi_j^+ + \frac{1}{2}\pi_j^e\right)\quad ...(3)$$

Now, the rank assigned to x_j is the sum of the expected value of U_{lj}, that is

$$r_j = W_j = \sum_{l=1}^{n} E(U_{ij}) = n\left(\pi_j^+ + \frac{1}{2}\pi_j^e\right)\quad ...(4)$$

Note that equations (1) – (4) enable all tied observations to be automatically assigned their mean ranks.

If x_j is not tied with any other observation in the sampled population, then $\pi_j^e = o,\ for\ some\ j = 1,2,...,n.$

Now

$\pi_j^+, \pi_j^e\ and\ \pi_j^o$ are respectively the proportions or the probabilities that a randomly selected observation x_j from the population is in fact not different from itself, or if different from itself, then that observation is higher (better, greater), the same as (equal to), or lower (worse, smaller) than all other observations in the sampled population. Their sampled estimates are respectively.

$$\hat{\Pi}_j^{\ +} = \frac{f_j^+}{n},\quad \hat{\Pi}_j^{\ e} = \frac{f_j^e}{n},\quad \hat{\Pi}_j^{\ 0} = \frac{f_j^o}{n}\quad ...(5)$$

Where $f_j^+, f_j^e\ and\ f_j^o$ are respectively the number of times the jth sampled observation x_j is either not different from itself, that that observation x_j is higher (better, greater), the same as (equal to) or lower (worse, smaller) than all the other sampled observations from the population X_1. In other words,

$f_j^+, f_j^e\ and\ f_j^o\ are\ respectively\ the\ total\ number\ of\ 1's,$

$\frac{1}{2}'s\ or\ 0's$ in the frequency distribution of the n values of these numbers in U_{lj}, for
l = 1,2 …, n; and some j = 1,2 ….n. Hence, the corresponding sample estimates of the ranks of x_j is from equation (4)

$$r_j = w_j = n\left(\hat{\Pi}_j^{\ +} + \tfrac{1}{2}\hat{\Pi}_j^{\ e}\right) = f_j^+ + \tfrac{1}{2} f_j^e \qquad \text{... (6)}$$

Now the sum of these ranks whether or not there are tied observations in a data is

$$R_1 = \sum_{j=1}^{n} r_j = \sum_{j=1}^{n}\left(f_j^+ + \frac{1}{2} f_j^e\right) = \frac{n(n+1)}{2} \qquad . . . (7)$$

However, the sums of squares of ranks are different depending on whether or not there are tied observations in the data and hence are assigned mean ranks. In general, especially when there are tied observations that is when all or some of the observations are tied and hence assigned mean ranks, the sum of squares of these ranks, that is the sum of squares of the ranks of the n sample observations drawn from population X_1 is

$$X_1^2 = \sum_{j=1}^{n}\left(f_j^+ + \frac{1}{2} f_j^e\right)^2$$

$$= \sum_{j=1}^{n} f_j^{+2} + \sum_{j=1}^{n}(f_j^e . f_j^+) + \frac{1}{4} f_j^{e2} \qquad . . . (8)$$

Note that in the absence of tied observations, that is when none of the observations are tied in scores so that none are assigned mean ranks, then
$\pi_j^e = 0, and\ f_j^e = 0\ for\ all\ j = 1,2, \ldots, n$

Then the resulting sum of squares of the ranks is

$$X_1^2 = \sum_{j=1}^{n} r_j^2 = \sum_{j=1}^{n} f_j^{+2} = \frac{n(n+1)(2n+1)}{6} \qquad . . . (9)$$

However, in the presence of ties in the data, equation 8 is the proper expression to use in obtaining a more efficient estimate of sum of squares of these ranks rather than equation 9, which results in an underestimate of the sum of squares or bias or underestimate of

$$B = \sum_{j=1}^{n} f_j^e\left(f_j^+ + \frac{1}{4} f_j^e\right) \qquad . . . (10)$$

Since f_j^e is now greater than zero
for j = 1, 2,… ,n.

Similarly, to obtain sample estimate of the sum of the products of the ranks assigned to n pairs of sampled observations drawn from populations X_1 and X_2 where x_{j1} is assigned the rank r_{j1} *and* x_{j2} is assigned the rank r_{j2} for j = 1, 2,…, n we have that

$$\mathrm{X}_{12} = \sum_{j=1}^{n} r_{j1} . r_{j2} = \sum_{j=1}^{n}(f_{j1}^+ + \frac{1}{2} f_{j1}^e)\left(f_{j2}^+ + \frac{1}{2} f_{j2}^e\right) \qquad . . . (11)$$

Which when expanded becomes.

$$\sum_{j=1}^{n} f_{j1}^+ f_{j2}^+ + \frac{1}{2}\sum_{j=i}^{n}(f_j^+ f_{j2}^+ + f_{j1}^e f_{j2}^+ + \frac{1}{2} f_{j1}^e f_{j2}^e) \qquad . . . (12)$$

Note that if there are no tied observations whatsoever in the data so that $f_{j1}^e = f_{j2}^e = 0$ for all
j = 1, 2,…n, then the sum of the products of these pairs of ranks become

$$S_{12} = \sum_{j=1}^{n} r_{j1} r_{j2} = \sum_{j=i}^{n} f_{j1}^+ . f_{j2}^+$$

$$= \frac{n(n+1)(2n+1)}{6} - \frac{\sum_{j=1}^{n} \mathrm{d}_j^2}{2} \qquad . . . (13)$$

where

$$d_j = f_{j1}^+ - f_{j2}^+ \qquad . . . (14)$$

is the difference between the pairs of ranks r_{j1} and r_{j2} assigned to the pairs of observations x_{j1} and x_{j2} drawn from populations X_1 and X_2 for j = 1, 2, …., n in the absence of tied observations in each of these populations. If only one of the two populations, X_2 say, has tied observations while there are no tied observations whatsoever in X_1 so that π_{j1}^e = 0 for j = 1, 2 …., n, then the resulting sample estimate or sum of products of the ranks of X_1 and X_2 is

$$X_{12} = \frac{n(n+1)(2n+1)}{6} - \frac{\sum_{j=1}^{n} \mathrm{d}_j^2}{2} + \frac{1}{2}\sum_{j=1}^{n} f_{j1}^+ f_{j2}^e \qquad . . . (15)$$

Since in this case, $f_{j1}^e = 0\ for\ all$
$j = 1, 2, \ldots, n\ while\ some$ or all f_{j2}^e are not equal to zero.

The above procedure would easily enable one systematically assign ranks to sampled observations drawn from a given population whether or not there are tied observations and simultaneously estimate the sum and mean of these ranks. Similarly, the method enables one to easily obtain more efficient estimates of the sum of squares and cross products of ranks of sample observations whether or not some or all of the observations are tied in values and hence assigned mean ranks. With these results, one may now proceed to estimate some ties adjusted statistics. For example, one may estimate ties adjusted Spearman rank correlation coefficient between pairs of observations drawn from populations X_1 and X_2.
This would yield

$$r_{xy} = r_{12} = \frac{\sum_{j=1}^{n} r_{j1} r_{j2} \sum_{j=1}^{n} r_{j1} \frac{\sum_{j=1}^{n} r_{j2}}{n}}{\sqrt{Var\,(r_{j1})\,Var\,(r_{j2})}} \qquad \text{... (16)}$$

Using ties adjusted sums of squares and product of ranks obtained above, in equation 16 and simplifying, we have

$$r_{xy} = r_{12} = \frac{\sum_{j=1}^{n}(f_{j1}^+ + \frac{1}{2} f_{j1}^e)\left(f_{j2}^+ + \frac{1}{2} f_{j2}^e\right) - \sum_{j=1}^{n}(f_{j1}^+ + \frac{1}{2} f_{j1}^e)\sum_{j=1}^{n}(f_{j2}^+ + \frac{1}{2} f_{j2}^e)/n}{\sqrt{\frac{\left(\sum_{j=1}^{n}(f_{j1}^+ + \frac{1}{2} f_{j1}^e)^2 - \sum_{j=1}^{n}(f_{j1}^+ + \frac{1}{2} f_{j1}^e)\right)^2}{n} \frac{(\sum_{j=1}^{n}(f_{j2}^+ + \frac{1}{2} f_{j2}^e)^2 - \sum_{j=1}^{n}(f_{j2}^+ + \frac{1}{2} f_{j2}^e))^2}{n}}}$$

. . . 17

Or when further simplified becomes

$$\mathrm{r}_{12} = \frac{\sum_{j=1}^{n} f_{j1}^+ f_{j2}^+ - \frac{1}{2}\sum_{j=1}^{n}(f_{j1}^+ . f_{j2}^e + f_{j1}^e f_{j2}^+ + \frac{1}{2} f_{j1}^e f_{j2}^e) - \frac{n(n+1)^2}{4}}{\sqrt{\left(\left(\sum_{j=1}^{n} f_{j1}^{+\,2} + \sum_{j=1}^{n} f_{j1}^e (f_{j1}^+ + \frac{1}{4} f_{j1}^e)\right) - -\frac{n(n+1)^2}{4})\left(\sum_{j=1}^{n} f_{j2}^{+\,2} + \sum_{j=1}^{n} f_{j2}^e f_{j2}^+ + \frac{1}{4} f_{j2}^e\right) - \frac{n(n+1)^2}{4})\right)}}$$

. . . (18)

Equation 18 is a general expression for a sample estimate of ties adjusted Spearman rank correlation coefficient between populations X_1 and X_2 whether or not there are tied observations is any or all the sampled populations. Equation 18 is a true and better estimate of spearman rank correlation coefficient between populations X_1 and X_2 when there are tied observations in the populations which are as a result of assigning mean ranks than its ties unadjusted alternative, namely (Siegel, 1956).

$$r_{xy} = r_{12} = \quad 1 - \frac{6\sum_{j=1}^{n} d_j^2}{n(n^2-1)} \qquad . . (19)$$

which is unadjusted for any possible ties between observations from the two populations X_1 and X_2. That is

$$r_{xy} = r_{12} \; = \frac{n(n^2-1)/12 \; - \sum_{j=1}^{n} d_j^2\,/2 \;\; + \frac{1}{2}\sum_{j=1}^{n}(f_{j1}^{+} \cdot f_{j2}^{e} + f_{j1}^{e} f_{j2}^{+} \; + \frac{1}{2}\; f_{j1}^{e} + f_{j2}^{e})}{n(n^2-1)/_{12} + \frac{1}{2}\sum_{j=1}^{n} f_{j1}^{e}\,(2f_{j1}^{+} + \frac{1}{2}\; f_{j1}^{e})]\; [n(n^2-1)/12 \; + \frac{1}{2}\sum_{j=1}^{n} f_{j2}^{e}\,(2f_{j2}^{+} + \frac{1}{2}\; f_{j2}^{e})]} \qquad . . . (20)$$

III. Illustrative Example

We now illustrate the proposed method with sample data on the scores in letter grades earned by a random sample of 13 students in two courses X_1 and X_2 taken during a semester.

Student Number	Score in X_1 (x_{j1})	Rank of x_{j1}	Score in X_2 (x_{j2})	Rank of x_{j2}	$d=x_{j1}-x_{j2}$	d^2
1	A^+	2	F	12	-10	100
2	A^+	2	F	12	-10	100
3	B^+	4.5	B^+	5	-0.5	0.25
4	D^-	9.5	A^-	3.5	6	36
5	D	9.5	F	12	-2.5	6.25
6	D^+	9.5	A	2	7.5	56.25
7	D	9.5	D^-	9	0.5	0.25
8	B	6	B^-	7	-1	1
9	D^-	9.5	D	9	0.5	0.25
10	F	13	A^+	1	12	144
11	D	9.5	A^-	3.5	6	36
12	A^+	2	D	9	-7	49
13	B^+	4.5	B	6	-1.5	2.25
						$\sum d_i^2$=531.50

Applying equation 1 to the student scores in the two courses, course 1, X_1 and course 2, X_2 we obtain values of U_{lj} and their summaries which are shown in tables 1 and 2 below.

***Table 1**: Values of U_{lj} (equation 1) for student scores in course 1 and other summary indices.*

S/N (*l*) Score in course X_1		**1 A^+**	**2 A^+**	**3 B^+**	**4 D^-**	**5 D**	**6 D^+**	**7 D**	**8 B**	**9 D^-**	**10 F**	**11 D**	**12 A^+**	**13 B^+**				
S/N (j)	**Score in course X_1**														f_{j1}^+	f_{j1}^o	f_{j1}^e	$r_{ji} = f_{j1}^+ \frac{1}{2} f_{j1}^e$
1	A^+	1	½	1	1	1	1	1	1	1	1	1	½	1	11	0	2	12
2	A^+	½	1	1	1	1	1	1	1	1	1	1	½	1	11	0	2	12
3	B^+	0	0	1	1	1	1	1	1	1	1	1	0	½	9	3	1	9.5
4	D^-	0	0	0	1	0	½	0	0	½	1	0	0	0	2	9	2	3
5	D	0	0	0	1	1	0	½	0	1	1	½	0	0	4	7	2	5
6	D^+	0	0	0	1	1	1	1	0	1	1	1	0	0	7	6	0	7
7	D	0	0	0	1	1	0	1	0	1	1	½	0	0	5	7	1	5.5
8	B	0	0	0	1	1	1	1	1	1	1	1	0	0	8	5	0	8
9	D^-	0	0	0	½	0	0	0	0	1	1	0	0	0	2	10	1	2.5
10	F	0	0	0	0	0	0	0	0	0	1	0	0	0	1	12	0	1
11	D	0	0	0	1	½	0	½	0	1	1	1	0	0	4	7	2	5
12	A^+	½	½	1	1	1	1	1	1	1	1	1	1	1	11	0	2	12
13	B^+	0	0	½	1	1	1	1	1	1	1	1	0	1	9	3	1	9.5
f_{lj}^+		1	1	4	11	9	7	8	6	11	13	8	1	4				
f_{lj}^O		10	10	8	1	3	5	3	7	1	0	3	10	8				
f_{lj}^e		2	2	1	1	1	1	2	0	1	0	2	2	1				
$r_{l1} = f_{l1}^+ + \frac{1}{2} f_{l1}^e$		2	2	4.5	11.5	9.5	7.5	9	6	11.5	13	9	2	4.5				

***Table 2:** Values of U_{lj} (equation 1) for student scores in course 2 and other summary indices*

S/N(*l*)		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	f_{j1}^+	f_{j1}^o	f_{j1}^e	$r_{j2} = f_{j2}^+ + \frac{1}{2} f_{j2}^e$
Score in course X_2		**F**	**F**	**B^+**	**A^-**	**F**	**A**	**D**	**B^-**	**D**	**A^+**	**A^-**	**D**	**B**				
S/N(j)	**Score in course X_2**																	
1	F	1	½	0	0	½	0	0	0	0	0	0	0	0	1	10	2	2
2	F	½	1	0	0	½	0	0	0	0	0	0	0	0	1	10	2	2
3	B^+	1	1	1	0	1	0	1	1	1	0	0	1	1	9	4	0	9
4	A^-	1	1	1	1	1	0	1	1	1	0	½	1	1	10	2	1	10.5
5	F	½	½	0	0	1	0	0	0	0	0	0	0	0	1	10	2	2
6	A	1	1	1	1	1	1	1	1	1	0	1	1	1	12	1	0	12
7	D	1	1	0	0	1	0	1	0	½	0	0	½	0	4	7	2	5
8	B^-	1	1	0	0	0	0	0	1	1	0	0	1	0	5	8	0	5
9	D	1	1	0	0	1	0	½	0	1	0	0	½	0	4	7	2	5
10	A^+	1	1	1	1	1	1	1	1	1	1	1	1	1	13	0	0	13
11	A^-	1	1	1	½	1	0	1	1	1	0	1	1	1	10	2	1	10.5
12	D	1	1	0	0	1	0	½	0	½	0	0	1	0	4	7	2	5
13	B	1	1	0	0	1	0	1	1	1	0	0	1	1	8	5	0	8
f_{l2}^+		11	11	5	3	10	2	7	7	8	1	3	8	6				
f_{l2}^O		0	0	8	9	1	11	4	6	3	12	9	3	7				
f_{l2}^e		2	2	0	1	2	0	2	0	2	0	1	2	0				
$r_{l1} = f_{l2}^+ + \frac{1}{2} f_{l2}^e$		12	12	5	3.5	11	2	8	7	9	1	3.5	9	6				

Table 3: *Estimation of Ties Adjusted Sum of Squares of ranks assigned to student scores in course 1*

S/N($_j$)	Score in course I (X_j)	Rank (r_{j1})	f_{j1}^{+}	f_{j1}^{e}	f_{j1}^{+2}	$f_{j1}^{+}.f_{j1}^{e}$	f_{j1}^{e2}	r_{j1}^{2}
1	A$^+$	12	11	2	121	22	4	144
2	A$^+$	12	11	2	121	22	4	144
3	B$^+$	9.5	9	1	81	9	1	90.25
4	D$^-$	3	2	2	4	4	4	9
5	D	5	4	2	16	8	4	25
6	D$^+$	7	7	0	49	0	0	49
7	D	5.5	5	1	25	5	1	30.25
8	B	8	8	0	64	0	0	64
9	D$^-$	2.5	2	1	4	2	1	6.25
10	F	1	1	0	1	0	0	1
11	D	5	4	2	16	8	4	25
12	A$^+$	12	11	2	121	22	4	144
13	B$^+$	9.5	9	1	81	9	1	90.25
Total					**704**	**111**	**28**	**822**

Table 4: *Estimation of Ties Adjusted Sum of Squares of Ranks assigned to student scores in course 2*

S/N(j)	Score in course I (X_j)	Rank (r_{j2})	f_{j2}^{+}	f_{j2}^{e}	f_{j2}^{+2}	$f_{j2}^{+}.f_{j2}^{e}$	f_{j2}^{e2}	r_{j2}^{2}
1	F	2	1	2	1	2	4	4
2	F	2	1	2	1	2	4	4
3	B$^+$	9	9	0	81	0	0	81
4	A$^-$	10.5	10	1	100	10	1	110.25
5	F	2	1	2	1	2	4	4
6	A	12	12	0	144	0	0	144
7	D	5	4	2	16	8	4	25
8	B$^-$	5	5	0	25	0	0	25
9	D	5	4	2	16	8	4	25
10	A$^+$	13	13	0	169	0	0	169
11	A$^-$	10.5	10	1	100	10	1	110.25
12	D	5	4	2	16	8	4	25
13	B	8	8	0	64	0	0	64
Total					**734**	**50**	**26**	**790.50**

The last column of table 1 shows values of r_{ji} (equation 6), the ties adjusted ranks assigned to student scores in course 1, ranked from the best grade, A$^+$, assigned the highest rank 13, to the lowest grade, F, assigned the rank 1 in the absence of ties and consistent with equation 1. Notice that the last row of table 1, $r_{i1}$are also the ranks that could be assigned to these grades if they are ranked from the lowest or worst score F assigned the rank 13, the highest rank, to the best or highest score A$^+$ assigned the rank 1 the lowest rank, in the absence of tied grades. Similar results obtained for student grades in course 2 are as shown in table 2.To obtain ties adjusted sums of squares of the ranks assigned to students' scores in course 1, we may proceed as in table 3. From this table's summaries, it is easily seen that the ties adjusted sum of squares of the ranks assigned to course 1 is

$$S_1^2 = \sum_{j=1}^{n} r_{j1}^2 = 822$$

$$= \sum_{j=1}^{13}(f_{j1}^{+} + \frac{1}{2} f_{j1}^{e})^2 = \sum_{j=1}^{13} f_{j1}^{+2} + \sum_{j=1}^{13}(f_{j1}^{+}.f_{j1}^{e}) + \frac{1}{4}\sum_{j=1}^{13} f_{j1}^{e2}$$

$$= 704 + 111 + \frac{1}{4}(28) = 822$$

Similarly, the ties adjusted sum of squares of ranks assigned to scores in course 2 is obtained from table 4 as

$$S_2^2 = \sum_{j=1}^{13} r_{j2}^2 = 790.50$$

$$= \sum_{j=1}^{13} f_{j2}^{+2} + \sum_{j=1}^{13}(f_{j2}^{+}.f_{j2}^{e}) + \frac{1}{4}\sum_{j=1}^{13} f_{j2}^{e2}$$

$$= 734 + 50 + \frac{1}{4}(26) = 790.5$$

The sample estimate of ties adjusted Spearman rank correlation coefficient between student scores in courses 1 and 2 may be obtained as shown in table 5.

***Table 5**: Estimate of Ties Adjusted Spearman Rank Correlation Coefficient between Student Scores in courses 1 and 2.*

S/N	Rank in course 1 (r_{j1})	Rank in course 2 (r_{j2})	Product of ranks ($r_{j1}r_{j2}$)	d_j^2	r_{j1}^2	r_{j2}^2	Difference ($d_j = r_{j1} - r_{j2}$)
1	12	2	24	100	144	4	10
2	12	2	24	100	144	4	10
3	9.5	9	85.5	0.25	90.25	81	0.5
4	3	10.5	31.5	56.25	9	110.25	-7.5
5	5	2	10	9	25	4	3
6	7	12	84	25	49	144	-5
7	5.5	5	27.5	0.25	30.25	25	0.5
8	8	5	40	9.0	64	25	3
9	2.5	5	12.5	6.25	6.25	25	-2.5
10	1	13	13.0	144	1	169	-12
11	5	10.5	52.5	30.25	25	110.25	-5.5
12	12	5	60	49	144	25	7
13	9.5	8	76	2.25	90.25	64	1.5
Total	**91**	**91**	**540.5**	**531.5**	**822**	**790.5**	

Using the summary values shown in table 5, we obtain the ties adjusted Spearman rank correlation coefficient between student scores in the two courses as from equation 16.

$$r_{12} = \frac{540.5 - (91)(91)/13}{\sqrt{(822 - 637)(790.5 - 637}} = -0.573$$

If we had ignored adjusting the ties for tied score and simply used the usual expression for calculating Spearman rank correlation coefficient, we would have obtained instead

$$r_{12} = | - \frac{6(531.5)}{13(13^2 - 1)} = -0.460$$

which is seen to be less, in absolute value, than – 0.573 obtained when adjustment had been made for tied scores by students in each course. It will be instructive to estimate the bias that would be introduced if adjustment were not made for tied scores by students in the courses. For example if in assigning ranks to scores by students in course 1 we had ignored assigning tied grade their mean ranks, we would have obtained the following ranking scheme for the grades as shown in table 6 below.

***Table 6**: Estimate of Bias due to Non-Adjustment for ties in the ranking of students' scores in course 1.*

S/N	Score in course (x_{j1})	Adjusted rank of score ($r_{j1}=r_{aj1}$)	Unadjusted Rank of Score (r_{ej1})	Difference ($d_{j1}= r_{aj1}- r_{ej1}$)	$r_{ej1}.\ d_{j1}$	d_{j1}^2	r_{ej1}^2
1	A$^+$	12	13	-1	-13	1	169
2	A$^+$	12	12	0	0	0	144
3	B$^+$	9.5	10	-0.5	-5	25	100
4	D$^-$	3	3	0	0	0	9
5	D	5	6	-1	-6	36	36
6	D$^+$	7	7	0	0	0	49
7	D	5.5	5	0.5	2.5	6.25	25
8	B	8	8	0	0	0	64
9	D$^-$	2.5	2	0.5	1	1	4
10	F	1	1	0	0	0	1
11	D	5	4	1	4	16	16
12	A$^+$	12	11	1	11	121	121
13	B$^+$	9.5	9	0.5	4.5	20.25	81
					-1.0	**226.5**	**819.0**

From summary data of table 6 we obtain the ties adjusted sum of squares of the ranks assigned to students scores in course 1 as

$$S_1^2 = \sum_{j=1}^{13} (r_{ej1} + d_{j1})^2 = \sum_{j=1}^{13} r_{ej1}^2 + 2\sum_{j=1}^{13} r_{ej1}\, d_{j1} + \sum_{j=1}^{13} d_{j1}^2$$
$$= 819 + 2(-1.0) + 226.5 = 1043.5$$

Here, adjustments for any tied observation in the data had been ignored and one had simply proceeded as if no observations are tied with one another. To do this we may proceed as follows. Suppose r_{ej1} is the rank that would have been expected to be assigned to x_{j1}, the sample observation drawn from population X_1 ignoring the possibility that this observation is tied with any other observation in X_1, and that $r_{aj1}=r_{j1}$ is the rank that is actually assigned to observation x_{j1} from population X_1, for j=1,2,…, n.

Let r_{ej2} *and* $r_{aj2} = r_{12}$ be similarly defined for $x_{j2,}$ the jth observation drawn from population X_2, for j = 1, 2,… n.

Let

$$d_{j1} = r_{ej1} - r_{aj1}; \quad d_{j2} = r_{ej2} - r_{aj2} \qquad . . . (20)$$

Being respectively the difference between the expected and actual ranks assigned to x_{j1} *and* x_{j2}, the jth observation drawn from population X_1 and X_2, then we would have that the ties adjusted sum of squares of the ranks assigned to observations from population X_1 say is

$$S_1^2 = \sum_{j=1}^{n} r_{aj1}^2 = \sum_{j=1}^{n} r_{j1}^2 = \sum_{j=1}^{n} (f_{j1}^+ + \frac{1}{2} f_{j1}^e)^2 = \sum_{j=1}^{n} (r_{ej1} + d_{j1})^2 \qquad . . . (21)$$

Or equivalently

$$S_1^2 = \sum_{j=1}^{n} r_{j1}^2 = \sum_{j=1}^{n} r_{ej1}^2 + 2\sum_{j=1}^{n} d_{j1}(r_{ej1} + \frac{1}{2} d_{j1})^2 = \frac{n(n+1)(2n+1)}{6} + 2\sum_{j=1}^{n} \mathrm{d_{j1}} \left(\mathrm{r_{ej1}} + \frac{1}{2} d_{j1}\right) \qquad . . . (22)$$

$$\text{Since} \sum_{j=1}^{n} r_{ej1}^2 = \frac{n(n+1)(2n+1)}{6}$$

A result that is valid since the sample observation x_{j1} , the jth observation from population X_1 is assumed not to be tied whatsoever with any other observation from population X_1, for all j= 1, 2,…, n.

Note that the bias that will arise from this assumption if it is not true is

$$\mathrm{b_1} = 2\sum_{j=1}^{n} d_{j1} (r_{ej1} + \frac{1}{2} d_{j1}) \qquad . . .(24)$$

which is zero if and only if d_{j1} = 0, for all j = 1,2,….n, that is if and only if there are no tied observations whatsoever in population X_1. Similarly the ties adjusted sum of the product of pairs of ranks assigned to pairs of sample observations drawn from populations X_1 and X_2 is estimated as

$$S_{12} = \sum_{j=1}^{n} (r_{aj1}.r_{aj2}) = \sum_{j=1}^{n} r_{j1}r_{j2} = \sum_{j=1}^{n} \left(f_{j1}^+ + \frac{1}{2} f_{j1}^e\right)(f_{j2}^+ + \frac{1}{2} f_{j2}^e) \qquad . . . (25)$$

Or equivalently

$$S_{12} = \sum_{j=1}^{n} (r_{ej1} + d_{j1})(r_{ej2} + d_{j2}) = \sum_{j=1}^{n} \mathrm{r_{ej1} r_{ej2}} + \sum_{j=1}^{n} (\mathrm{r_{ej1} d_{j2}} + r_{ej2} d_{j1} + d_{j1} d_{j2}) \qquad . . (26)$$

Or equivalently, letting

$$d_{ej} = r_{ej1} - r_{ej2} \qquad . . . (27)$$

We have

$$S_{1j} = \frac{n(n+1)(2n+1)}{6} + \sum_{j=1}^{n} \frac{d_{ej}^2}{2} + \sum_{j=1}^{n} (r_{ej1} d_{j2} + r_{ej2} d_{j1} + d_{j2}) \qquad . . . (28)$$

Note that if any of the sampled populations X_1 say, has no tied observations so that d_{j1} = 0 for all j = 1,2,….., n then the ties adjusted sum of products of these pair of ranks is estimated as

$$S_{12} = \frac{n(n+1)(2n+1)}{6} - \sum_{j=1}^{n} \frac{d_{ej}^2}{2} + \sum_{j=1}^{n} r_{ej1}.d_{j2} \qquad . . . (23)$$

It is easily shown using these results that the ties adjusted sample estimate of the spearman rank correlation coefficient between populations X_1 and X_2 is

$$\mathrm{r_{12}} = \frac{\frac{n(n^2-1)}{12} - \sum_{i=1}^{n} \frac{d_{ej}^2}{2} + \sum_{j=1}^{n} (r_{ej1}.d_{j2} + r_{ej2}.d_{j1} + d_{j1}.d_{j2})}{\sqrt{\left\{\frac{[n(n^2-1)}{12} + 2\sum_{i=1}^{n} d_{j1}(r_{ej1} + \frac{1}{2} d_{j1})][\frac{n(n^2-1)}{12} + 2\sum_{j=1}^{n} d_{j2}(r_{ej2} + \frac{1}{2} d_{j2})\right\}}} \qquad . . . (30)$$

Note that equation 30 is the same as the usual sample estimate of Spearman rank correlation coefficient between observations from population X_1 and X_2 if any and only if there is not tied observations whatsoever in any of the two populations in which case $d_{j1} = d_{j2} = 0$, so that we have

$$r_{12} = 1 - \quad r_{12} = 1 - \frac{\sum_{i=1}^{n} \mathrm{d_j^2}/2}{\mathrm{n(n^2-1)}} \qquad . . . (31)$$

as can easily be shown from equation 30.

REFERENCES

[1] Gibbons J.D. (1973). Nonparametric Statistical Inference. McGrawhill Book Company, New York.

[2] Hollander, M and Wolfe D.A. (1999). Nonparametric Statistical Methods (2nd Edition). Wiley – Inter Science ,New York.

[3] Oyeka C.A., Osuji, GA and Nwankwo, C.C. (2013) “Simplified Method of Rank Determination in Rank Order Statistics” Mathematical Theory and Modeling Vol 3, No 14.

[4] Oyeka Ikewelugo Cyprian Anene, Nwankwo Chike H, Awopeju K. Abidemi (2014). The G- Method for Rank Determination in Rank Order Statistics for ‘C Samples. American Journals of Theoretical and Applied Statistics. Vol 3, No1, pp 18-24.

[5] Siegel Sidney (1956): “Nonparametric Statistics For the Behavioural Sciences”. International Students Edition. McGraw – Hill, KOGAKUSHA Ltd.