Computational Journal of Mathematical and Statistical Sciences
2(2), 240–250
DOI:10.21608/CJMSS.2023.229746.1015
https://cjmss.journals.ekb.eg/

A new Method for Computing and TestingThe significance of the Spearman Rank Correlation

Mahmoud Eltehiwy[1*], **Abu Bakr Abdul-Motaal**[2]

[1] statistics department, the Politics and Economics faculty, Beni -Suef university, Egypt; mahmoudtehiwy@gmail.com

[2] Faculty of Commerce, Department of Statistics, South Valley University, Egypt; abubakrabdulmotaal@gmail.com

* **Correspondence:** mahmoudtehiwy@gmail.com

Abstract: To compute and test the significance of the Spearman rank correlation coefficient, the differences between ranks are used, and the Spearman correlation coefficient is used as a test statistic. This study has introduced an efficient rank correlation formula like the method earlier derived by Spearman. The formula is based on a new statistic that depends on the sum of the ranks rather than their difference, as in Spearman's formula. The formula was derived and tested with real data. Hence, if there are no ties in the data, the result shows that the formula gives the same result as Spearman's formula. The formula is simple to use and does not include a negative sign during calculation. we clarify the relationship between the new test statistic and the Spearman rank correlation coefficient and establish the exact distribution of the new test statistic using the exact distribution of the Spearman rank correlation coefficient. When there are no ties, it is advised to follow this formula.

Keywords: Asymptotic normality, exact distribution, rank correlation coefficient, tied observations, significance.

Mathematics Subject Classification: 70K45; 62H20; 11K06.

Received: 16 August 2023; **Revised:** 1 September 2023; **Accepted:** 21 September 2023; **Online:** 24 September 2023.

1. Introduction

One of the earliest applications of non-parametric statistics that most social scientists have learned about is Spearman's rho. It is widely available in most statistical software packages, has an intuitive conception as the Pearson correlation between two ranking variables, and offers an immediate replacement when the distributional assumption of normality is questioned.

The rank correlation coefficient as proposed by [1], when there are no ties, is defined as follows:

$$r_s = 1 - \frac{6 \sum_{i=1}^{n} D_i^2}{n\left(n^2 - 1\right)}, \tag{1.1}$$

where $D_i^2 = [R(X_i) - R(Y_i)]^2$ are the squared rank differences between variables X and Y and n is the total number of measured units (i.e., the sample size). This is the most commonly used computational formula to obtain the Spearman correlation and perhaps the most easily recognized one among introductory statistics textbooks in the social sciences. Its standard error and sampling distribution were extensively studied by [2], who found that, for the null case, $1/\sqrt{n-3}$ and the normal distribution are respectively good approximations as the sample size grows arbitrarily large. For the non-null case, most work has relied on computer simulations, which can be reviewed in [3, 4, 5].

In rank correlation analysis, it is common to test the null hypothesis that there is no correlation in the population between the paired ranks. There are numerous tables of critical values for r_s, and if r_s is greater than the relevant critical value, the null hypothesis is rejected. The "Hotelling-Pabst test" used $\sum_{i=1}^{n} D_i^2$ as the test statistic for rank correlation testing rather than r_s. (See [6]). Critical values for various sample sizes, n, and levels of significance are provided in published tables. The most comprehensive r_s tables are those of [7, 8].

Tables of the exact distribution of r_s and $\sum D^2$ are available in the literature. The table of the exact quantiles of r_s for sample sizes 4 through 30 was prepared by [9, 10], who also gave tables of the probability function of $\sum D^2$.

The purpose of this paper is to develop an alternative approach for computing and testing the significance of the rank correlation coefficient. We used the sum of ranks instead of the differences between ranks for this purpose, and a new test statistic is introduced. if there are no ties in the data, the result shows that the new formula based on the new statistics gives the same result as Spearman's formula. The formula is simple to use and does not include a negative sign during calculation. We show the relationship between the new test statistic and r_s. We construct the exact distribution of the new test statistic using the exact distribution of r_s. The significance of r_s is tested using the critical values that we will get for the new statistics in a numerical example.

The notion on which the proposed approach is based is the following: for a set of n pairs of $R(X)$ and $R(Y)$, let T be the sum of the ranks of the two random variables X and Y as follows:

$$T_i = R(X_i) + R(Y_i).$$

When two rankings of X and Y are identical, the situation in which r_s reaches its maximum value 1, $\sigma^2(T)$, the variance of T, will take on its maximum value $\frac{(n^2-1)}{3}$. This is because the totals of ranks in this case would be some permutations of $2, 4, 6, \ldots, 2n$. On the other hand, when one ranking is the reverse of the other, in the situation in which r_s reaches its minimum value of -1, the variance $\sigma^2(T)$ will take on its minimum value of zero (in this case, $T_i = n+1$ for $i = 1, 2, \ldots, n$). That is, $\sigma^2(T)$ varies from zero to $\frac{(n^2-1)}{3}$. Zero signifies perfect negative disagreement, whereas 1 signifies perfect positive agreement. This may indicate, as a point of interest, that a relationship between $\sigma^2(T)$ and r_s might be investigated. In other words, the rank correlation coefficient r_s may be defined in terms of the value of $\sigma^2(T)$ which, in turn, can be used as an indicator of the strength of association between X and Y (as measured by r_s). Hence, the totals of ranks, rather than the differences between ranks are to be considered.

For example when $n = 5$, we can obtain the probability distribution of T as follows:

T	2	3	4	5	6	7	8	9	10
g (T)	1/25	2/25	3/25	4/25	5/25	4/25	3/25	2/25	1/25

So, we can conclude that the distribution of T (the sum of all possible sets of pairings of $R(X)$ and $R(Y)$) is defined as

$$g(T) = \begin{cases} \frac{T-1}{n^2} & T = 2, 3, \ldots, n+1 \\ \frac{(2n+1)-T}{n^2} & T = n+2,\ n+3, \ldots, 2n \end{cases}, \tag{1.2}$$

and hence

$$E(T) = \sum_{i=1}^{n} \frac{T_i(T_i - 1)}{n^2} + \sum_{i=n+1}^{2n-1} \frac{T_i[(2n+1) - T_i]}{n^2}.$$

$$E(T) = \frac{2n(n+1)(n+2)}{6n^2} + \frac{4n(n+1)(n-1)}{6n^2}$$

It follows after some simplification that

$$E(T) = n + 1. \tag{1.3}$$

Similarly,

$$E\left(T^2\right) = \sum_{i=1}^{n} \frac{T_i^2(T_i - 1)}{n^2} + \sum_{i=n+1}^{2n-1} \frac{T_i^2[(2n+1) - T_i]}{n^2},$$

$$E\left(T^2\right) = \frac{n(n+1)(n+2)(3n+5)}{12n^2} + \frac{n(n+1)(11n+10)(n-1)}{12n^2}$$

which leads to

$$E\left(T^2\right) = \frac{(n+1)(7n+5)}{6}. \tag{1.4}$$

It follows from (1.3) and (1.4) that

$$Var(T) = \frac{n^2 - 1}{6}. \tag{1.5}$$

It should be noted that $\sigma^2(T)$ concerns one set of pairings of $R(X)$ and $R(Y)$, whereas $Var(T)$ concerns all possible different sets of pairings of $R(X)$ and $R(Y)$.

This paper is organized as follows. In Section 2, the relationship between r_s and $\sigma^2(T)$ is presented. Section 3 described the exact distribution of $\sigma^2(T)$. An alternative way for obtaining the mean and variance of $\sigma^2(T)$ is discussed in Section 4. Section 5 provides tests for the significance of r_s for small and large samples. The conclusion will be covered in Section 6.

2. The relationship between r_s and $\sigma^2(T)$

For a given set of untied rankings for X and Y, a possible relationship between r_s and $\sigma^2(T)$ can be investigated in the following manner:
The variance $\sigma^2(T)$ can be expressed in terms of $R(X)$ and $R(Y)$ as follows:

$$\sigma^2(T) = \frac{1}{n}\sum_{i=1}^{n} T_i^2 - \frac{1}{n^2}\left(\sum_{i=1}^{n} T_i\right)^2,$$

$$= \frac{1}{n}\sum_{i=1}^{n}[R(X_i)+R(Y_i)]^2 - \frac{1}{n^2}\left\{\sum_{i=1}^{n}[R(X_i)+R(Y_i)]\right\}^2.$$

Since $\sum_{i=1}^{n} R^2(X_i) = \sum_{i=1}^{n} R^2(Y_i) = \frac{n(n+1)(2n+1)}{6}$,
then

$$\sigma^2(T) = \frac{2}{n}\sum_{i=1}^{n} R(X_i)R(Y_i) - \frac{(n+1)(n+2)}{3}, \tag{2.1}$$

and since $\sum_{i=1}^{n} D_i^2 = \sum_{i=1}^{n}[R(X_i) - R(Y_i)]^2$,

$$= \sum_{i=1}^{n}\left\{R^2(X_i) + R^2(Y_i)\right\} - 2\sum_{i=1}^{n} R(X_i)R(Y_i),$$

then

$$\sum_{i=1}^{n} D_i^2 = \frac{n(n+1)(2n+1)}{3} - 2\sum_{i=1}^{n} R(X_i)R(Y_i). \tag{2.2}$$

From (2.1) and (2.2) we find

$$\sum_{i=1}^{n} D_i^2 = \frac{n\left(n^2-1\right)}{3} - n\sigma^2(T). \tag{2.3}$$

Now, substituting equation (2.3) into equation (1.1), we get

$$r_s = \left(\frac{6}{n^2-1}\right)\sigma^2(T) - 1. \tag{2.4}$$

3. The Exact distribution of $\sigma^2(T)$

The exact distribution of $\sigma^2(T)$ can be obtained by enumerating all possible permutations of the n untied ranks for the two variables. This provides $n!$ distinct sets of pairings of $R(X)$ and $R(Y)$, and hence provides a set of $n!$ values for $\sigma^2(T)$. These values can be used to obtain the exact distribution of $\sigma^2(T)$. The $n!$ distinct sets of pairings of $R(X)$ and $R(Y)$ are determined by keeping the $R(X)'s$ fixed and permuting the $R(Y)'s$ with equal probabilities $(n!)^{-1}$. The variance $\sigma^2(T)$ can then be found for each pair of $R(X)$ and $R(Y)$. However, it is much easier to obtain the exact distribution of $\sigma^2(T)$ in this case by using its relationship with r_s whose exact distribution is known. Tables of the exact distribution of r_s and $\sum D^2$ are available in the literature. The table of the exact quantiles of r_s for sample sizes 4 through 30 was prepared by [9, 10], which gives tables of the probability function of $\sum D^2$. In this paper, the exact distribution of $\sigma^2(T)$ is constructed (Table 1) using the exact distribution of r_s and the relationship between $\sigma^2(T)$ and r_s as defined in (2.4). Table 1 shows the values of $\sigma^2(T)$ for selected quantiles, under the assumption of independence, for sample sizes ranging from 4 to 30.

To obtain the corresponding lower quantiles for $\sigma^2(T)$, let us first consider the mean and variance of $\sigma^2(T)$. Since the mean and variance of r_s (see [11]) are given by:

$$E(r_s) = 0 \;\; \& \;\; Var(r_s) = \frac{1}{n-1} \tag{3.1}$$

Table 1. Exact quantiles for $\sigma^2(T)$

n	P=0.900	P=0.950	P=0.975	0.990	0.995	0.999
4	4.500	4.500				
5	6.800	7.200	7.600	7.600		
6	9.333	10.333	10.667	11.000	11.334	
7	12.286	13.429	13.960	14.857	15.143	15.714
8	15.750	17.000	18.000	19.000	19.500	20.250
9	19.556	21.111	22.444	23.556	24.223	25.333
10	23.800	25.600	27.001	28.559	29.400	30.801
11	28.364	30.546	32.182	34.000	34.910	36.728
12	33.333	35.667	37.666	39.833	41.167	43.334
13	38.615	41.384	43.537	46.001	47.538	50.145
14	44.285	47.427	49.858	52.715	54.428	57.427
15	50.400	53.868	56.668	59.733	61.734	65.199
16	56.873	60.626	63.750	67.252	69.377	73.376
17	63.648	67.766	71.294	75.058	77.530	81.998
18	70.780	75.334	79.221	83.334	85.999	91.000
19	78.420	83.370	87.474	91.998	94.950	100.422
20	86.297	91.697	96.099	101.100	104.299	110.297
21	94.666	100.379	105.241	110.572	113.997	120.670
22	103.273	109.456	114.640	120.452	124.179	131.360
23	112.350	118.958	124.520	130.698	134.693	142.437
24	121.747	128.752	134.751	141.335	145.667	154.004
25	131.518	138.965	145.361	152.402	157.040	166.005
26	141.615	149.614	156.307	163.845	168.773	178.380
27	152.152	160.597	167.707	175.703	180.969	191.185
28	162.994	171.934	179.424	187.933	193.505	204.363
29	174.202	183.582	191.590	200.480	206.416	217.938
30	185.793	195.667	204.073	213.527	219.731	231.927

The mean and variance of $\sigma^2(T)$ are defined by taking the means and variances on both sides of (2.4), and then by (3.1) we find:

$$E\left[\sigma^2(T)\right] = \frac{n^2-1}{6}, \tag{3.2}$$

and

$$Var\left[\sigma^2(T)\right] = \frac{(n-1)(n+1)^2}{36}. \tag{3.3}$$

Note: the corresponding lower quantiles are obtained from the following equation:

$$\sigma^2_{1-P}(T) = \frac{n^2-1}{3} - \sigma^2_P(T).$$

In view of the symmetry of the distribution of r_s about zero with variance $\frac{1}{n-1}$, it follows from (2.4) that the distribution of $\sigma^2(T)$ is symmetric about $\frac{(n^2-1)}{6}$ with the variance defined in (3.3). Consequently, for $r_s = 0$, the value of $\sigma^2(T)$ will be equal to $\frac{(n^2-1)}{6}$. As a result, the closer the value of $\sigma^2(T)$ is to $\frac{(n^2-1)}{6}$, the weaker the association between X and Y (as measured by r_s), and vice versa.

Now, since the distribution of $\sigma^2(T)$ is symmetric about $\frac{(n^2-1)}{6}$, the corresponding lower quartile $\sigma^2_{1-P}(T)$ for a given upper quantile $\sigma^2_P(T)$ can then be determined by using the folowing formula:

$$\sigma^2_{1-P}(T) = \frac{n^2-1}{3} - \sigma^2_P(T) \tag{3.4}$$

For example: for $n = 24$, the 0.01 and 0.1 quantiles for $\sigma^2(T)$ are given as follows:

$$\sigma^2_{0.01}(T) = \frac{(24)^2-1}{3} - 141.335 = 50.332$$

Similarly,

$$\sigma^2_{0.1}(T) = 191.667 - 121.747 = 69.92$$

4. An alternative way for obtaining the mean and variance of $\sigma^2(T)$

Since $\sigma^2(T)$ is given by

$$\sigma^2(T) = \frac{1}{n}\sum_{i=1}^{n} T_i^2 - \frac{1}{n^2}\left(\sum_{i=1}^{n} T_i\right)^2,$$

then, the mean and variance of $\sigma^2(T)$ can be defined using the mean and variance of $\sum_{i=1}^{n} T_i$ and $\sum_{i=1}^{n} T_i^2$. To define the mean of $\sigma^2(T)$, the distribution of T as given in (1.2) can be used, and hence the $n!$ distinct sets of pairings of $R(X)$ and $R(Y)$ should be considered.

Let $Z_j = \sum_{i=1}^{n} T_{ij}$ and $S_j = \sum_{i=1}^{n} T_{ij}^2$ for $j = 1, 2, 3, \ldots, n!$,

where T_{ij} represents the ith observation for T ($i = 1, 2, \ldots, n$) on the jth set ($j = 1, 2, \ldots, n!$).

Now,

$$E\left(\sum_{i=1}^{n} T_i\right) = \frac{\sum_{j=1}^{n!} Z_j}{n!} = \sum_{i=1}^{n}\sum_{j=1}^{n!} \frac{T_{ij}}{n!}. \tag{4.1}$$

Since $E(T) = \sum_{i=1}^{n}\sum_{j=1}^{n!} \frac{T_{ij}}{n(n!)}$, then, by (1.3) and (4.1), we find:

$$E\left(\sum_{i=1}^{n} T_i\right) = n(n+1), \tag{4.2}$$

or

$$E\left(\sum_{i=1}^{n} T_i\right) = E\left[\sum_{i=1}^{n} R(X_i)\right] + E\left[\sum_{i=1}^{n} R(Y_i)\right] = E\left[\frac{n(n+1)}{2} + \frac{n(n+1)}{2}\right] = n(n+1).$$

$$E\left(\sum_{i=1}^{n} T_i^2\right) = \frac{\sum_{j=1}^{n!} S_j}{n!} = \sum_{i=1}^{n}\sum_{j=1}^{n!} \frac{T_{ij}^2}{n!}, \tag{4.3}$$

Since $E\left(T^2\right) = \sum_{i=1}^{n} \sum_{j=1}^{n!} \frac{T_{ij}^2}{n(n!)}$, then, by (1.4) and (4.3), we get:

$$E\left(\sum_{i=1}^{n} T_i^2\right) = \frac{n(n+1)(7n+5)}{6} \tag{4.4}$$

For example, if $n = 3$, then, there are 3! distinct sets of pairings of $R(X)$ and $R(Y)$ are determined by keeping the $R(X)^{'s}$ fixed and permuting the $R(Y)^{'s}$ as follows:

$R(X_i)$	1	2	3
$R(Y_i)$	1	2	3
T_{i1}	2	4	6

$$Z_1 = \sum_{i=1}^{3} T_{i1} = 12, S_1 = \sum_{i=1}^{3} T_{i1}{}^2 = (4 + 16 + 36) = 56.$$

$R(X_i)$	1	2	3
$R(Y_i)$	1	3	2
T_{i2}	2	5	5

$$Z_2 = \sum_{i=1}^{3} T_{i2} = 12, S_2 = \sum_{i=1}^{3} T_{i2}{}^2 = (4 + 25 + 25) = 54.$$

$R(X_i)$	1	2	3
$R(Y_i)$	2	1	3
T_{i3}	3	3	6

$$Z_3 = \sum_{i=1}^{3} T_{i3} = 12, S_3 = \sum_{i=1}^{3} T_{i3}{}^2 = (9 + 9 + 36) = 54.$$

$R(X_i)$	1	2	3
$R(Y_i)$	2	3	1
T_{i4}	3	5	4

$$Z_4 = \sum_{i=1}^{3} T_{i4} = 12, S_4 = \sum_{i=1}^{3} T_{i4}{}^2 = (9 + 25 + 16) = 50.$$

$R(X_i)$	1	2	3
$R(Y_i)$	3	1	2
T_{i5}	4	3	5

$$Z_5 = \sum_{i=1}^{3} T_{i5} = 12, S_5 = \sum_{i=1}^{3} T_{i5}{}^2 = (16 + 9 + 25) = 50.$$

$R(X_i)$	1	2	3
$R(Y_i)$	3	2	1
T_{i6}	4	4	4

$$Z_6 = \sum_{i=1}^{3} T_{i6} = 12, S_6 = \sum_{i=1}^{3} T_{i6}{}^2 = (16+16+16) = 48.$$

Then,

$$E\left(\sum_{i=1}^{n=3} T_i\right) = \frac{\sum_{j=1}^{3!} Z_j}{3!} = \sum_{i=1}^{3}\sum_{j=1}^{3!} \frac{T_{ij}}{3!} = 12.$$

or

$$E\left(\sum_{i=1}^{n=3} T_i\right) = 3(3+1) = 12.$$

$$E\left(\sum_{i=1}^{3} T_i^2\right) = \frac{\sum_{j=1}^{3!} S_j}{3!} = \sum_{i=1}^{3}\sum_{j=1}^{3!} \frac{T_{ij}^2}{3!} = \frac{(56+54+54+50+50+48)}{6} = 52.$$

Or

$$E\left(\sum_{i=1}^{3} T_i^2\right) = \frac{3(4)(26)}{6} = 52.$$

It can also be proven that $Var\left(\sum_{i=1}^{n} T_i^2\right) = Var\left(\sum_{i=1}^{n} D_i^2\right)$,
Since $Var\left(\sum_{i=1}^{n} D_i^2\right) = \frac{n^2(n-1)(n+1)^2}{36}$, (See [12]).
It follows that

$$Var\left(\sum_{i=1}^{n} T_i^2\right) = \frac{n^2(n-1)(n+1)^2}{36} \tag{4.5}$$

Now, we are prepared to find the mean and variance of $\sigma^2(T)$

$$E\left[\sigma^2(T)\right] = \frac{1}{n}E\left(\sum_{i=1}^{n} T_i^2\right) - \frac{1}{n^2}E\left(\sum_{i=1}^{n} T_i\right)^2.$$

From (4.2) and (4.4) we get

$$E\left[\sigma^2(T)\right] = \frac{1}{n}\left[\frac{n(n+1)(7n+5)}{6}\right] - (n+1)^2.$$

It follows after some simplification that

$$E\left[\sigma^2(T)\right] = \frac{n^2-1}{6}.$$

$$Var\left[\sigma^2(T)\right] = Var\left[\frac{\sum_{i=1}^{n} T_i^2}{n} - \frac{\left(\sum_{i=1}^{n} T_i\right)^2}{n^2}\right] = Var\left(\frac{\sum_{i=1}^{n} T_i^2}{n}\right) + Var\left(\frac{\left(\sum_{i=1}^{n} T_i\right)^2}{n^2}\right).$$

Since, $\frac{\left(\sum_{i=1}^{n} T_i\right)^2}{n^2}$ is a constant, then,
$Var\left[\sigma^2(T)\right] = Var\left(\frac{\sum_{i=1}^{n} T_i^2}{n}\right)$, and by (4.5), we get

$$Var\left[\sigma^2(T)\right] = \frac{(n-1)(n+1)^2}{36},$$

which agrees with the results obtained before in (3.2) and (3.3).

5. Testing the significance of r_s

A common desire in rank correlation analysis is to test the null hypothesis that there is no correlation in the population between the paired ranks, i.e., we wish to test the two-tailed hypotheses $H_0 : r_s = 0$ *vs*. $H_1 : r_s \neq 0$. There are many tables of critical values for r_s, and if r_s is greater than the relevant critical value, then H_0 is rejected. The use of $\sum_{i=1}^{n} D_i^2$ instead of r_s as the test statistic for rank correlation testing is sometimes called the "Hotelling-Pabst test". (see [6]). $\sum_{i=1}^{n} D_i^2$ is small when r_s is large, and H_0 is rejected if $\sum_{i=1}^{n} D_i^2$ is less than the critical value. Published tables offer critical values for various sample sizes, n, and levels of significance, α. The most extensive of such tables for r_s are those of [8] and, with slight improvements, of [7].

5.1. Small Samples

For small samples ($n \leq 30$), the exact probability distribution of $\sigma^2(T)$ given in Table 1 is used to test the significance of r_s. That is , $\sigma^2(T)$ is used as a test statistic to test for independence between X and Y. The hypotheses take the form:
A. $H_0 : r_s = 0$ versus $H_1 : r_s \neq 0$
B. $H_0 : r_s = 0$ versus $H_1 : r_s > 0$
C. $H_0 : r_s = 0$ versus $H_1 : r_s < 0$

For a given significance level α, and using Table 1, the H_0 in B is rejected if the observed value of $\sigma^2(T)$ exceeds the $1 - \alpha$ quantile. The H_0 in C is rejected if the observed value $\sigma^2(T)$ is smaller than the α quantile, and the H_0 in A is rejected if the observed value of $\sigma^2(T)$ exceeds the $1 - \alpha/2$ quantile or if $\sigma^2(T)$ is less than the $\alpha/2$ quantile.

Example: Apply the preceding hypotheses (A,B, and C) to the following rankings for X and Y, using $\alpha = 0.05$.

$R(X_i)$	3	6	8	1	4	7	2	5	9
$R(Y_i)$	1	8	9	3	2	6	4	7	5

The values of T_i in this example are: 4, 14, 17, 4, 6, 13, 6, 12, and 14. Thus, $\sigma^2(T)$ equals 22. From table 1, and for $n = 9$, we find:
The 0.95 and 0.975 quantiles for $\sigma^2(T)$ are 21.111 and 22.444 respectively. The corresponding lower quantiles for $\sigma^2(T)$ can be obtained using formula (3.4). That is, the 0.05 and 0.025 quantiles for $\sigma^2(T)$ are, respectively, $(26.667 - 21.111 = 5.556)$ and $(26.667 - 22.444 = 4.223)$.

Since the observed value of 22 for $\sigma^2(T)$ is larger than the value of the 0.95 quantile for $\sigma^2(T)$, then the H_0 in B is rejected at the significance level of 0.05. The H_0 in C is accepted at the significance level of 0.05 because the observed value of $\sigma^2(T)$ is larger than the 0.05 quantile for $\sigma^2(T)$. Finally, the H_0 in A is accepted at the significance level of 0.05 because the observed value of $\sigma^2(T)$ lies between the two values of the 0.025 and 0.975 quantiles (note that: $4.223 < 22 < 22.444$).

It should be pointed out that the test statistic $\sigma^2(T)$ gives results as if the r_s test statistic was used.

5.2. Large Samples:

When the sample size is greater than 30, we cannot use table 1 to test the significance of r_s. In this case, it is probably accurate enough to use the normal approximation of the $\sigma^2(T)$ distribution.

For $n > 30$, $r_s\sqrt{n-1}$ is distributed approximately as the standard normal distribution (See [10]). As a result of (2.4), it follows that

$$Z = \frac{\sigma^2(T) - \frac{(n^2-1)}{6}}{\sqrt{\frac{(n-1)(n+1)^2}{36}}} \tag{5.1}$$

is also distributed as the standard normal. That is, testing $\sigma^2(T) = \frac{(n^2-1)}{6}$ is equivalent to testing $r_s = 0$. Thus, the associated probability under H_0 of any value as extreme as an observed value of $\sigma^2(T)$ may be determined by computing the Z associated with that value using formula (5.1) and then determining the significance of that Z by referring to the table of the standard normal distribution.

Example: Suppose we have: $n = 40$ and $\sigma^2(T) = 375$.

Test: $H_0 : r_s = 0$ versus $H_1 : r_s > 0$, using $\alpha = 0.01$.

In this example, using formula (5.1) we get $Z = 2.542$. From the table of the standard normal distribution, we find the significance level associated with 2.542 is $\widehat{\alpha} = 0.0039$. Since $\widehat{\alpha} < 0.01$, then H_0 is rejected at $\alpha = 0.01$. Alternatively, from the table of the standard normal distribution, we find $Z_{0.99} = 2.326$. Since $2.542 > Z_{0.99}$, we can reject H_0 at $\alpha = 0.01$.

6. Conclusions

In this paper, we developed an alternative approach for computing and testing the significance of the rank correlation coefficient. To compute and test the significance of the Spearman rank correlation coefficient, we used the totals of ranks, instead of the differences between ranks for this purpose, and a new test statistic is introduced. we showed the relationship between the new test statistic and r_s. We constructed the exact distribution of the new test statistic (Table 1) using the exact distribution of r_s. The new test statistic is used in testing the significance of the rank correlation coefficient and gives identical results to those obtained by the Spearman correlation coefficient test statistic.

Finally, in the next work, it is aimed to develop the new statistic in nonparametric correlation analysis so it can both negate the impact of ties as well as offer a closer approximation to the parametric Karl-Pearson's product-moment correlation coefficient. In addition, we will demonstrate that the new rank correlation coefficient formula has a direct link with Kendall's coefficient of concordance.

Conflict of interest: The authors state that they have no financial or other conflicts of interest to disclose with connection to this research.

References

1. Spearman, C. (1904). The proof and measurement of association between two things. Am. J. Psychol. 15, 72-101.
2. Fieller, E. C., Hartley, H. O., and Pearson, E. S. (1957). Tests for rank correlation coefficients. Biometrika, 44 ,470–481. doi:10.2307/2332878.
3. Caruso, J. C., and Cliff, N. (1997). Empirical size, coverage, and power of confidence intervals for Spearman's rho. Educ. Psychol. Meas. 57, 637–654. doi:10.1177/0013164497057004009.

4. David, F. N., and Mallows, C. L. (1961). The variance of Spearman's rho in normal samples. Biometrika, 48, 19–28. doi:10.2307/2333126.
5. Zimmerman, D. W. Zumbo, B. D., and Williams, R. H. (2003). Bias in estimation and hypothesis testing of correlation. Psicologica, 24(1), 133–158. Retrieved from https://dialnet.unirioja.es/servlet/ articulo? codigo=647711
6. Hotelling, H., and Pabst, M. R. (1936). Rank correlation and tests of significance involving no assumption of normality, *Ann. Math. Stat.* **7 (1)** , 29–43.
7. Ramsey, P.H. (1988). Critical values for Spearman's rank order correlation, J. Educ. Stat. 14 (3), 245–253.
8. Zar, J. H. (1996). *Biostatistical Analysis*, 3rd Ed. Prentice-Hall, Upper Saddle River.
9. Glasser, G. J., and Winter, R. F. (1961). Critical Values of the coefficient of rank correlation for testing hypothesis of independence. Biometrika, 48, 444-448.
10. Kendall, M., and Gibbons, J. D. (1990). *Rank Correlation Methods*, 5^{th} ed., Edward Arnold, London.
11. Gibbons, J. D. (1971). *Nonparametric Statistics Inference*, McGraw-Hill, New York.
12. Lehman, E. L., and D'Abrera, M. H. S. (1975). *Nonparametric*, McGraw-Hill. New York.

SASAR
Open Access Journal