Modified Wilcoxon Signed-Rank Test

Ikewelugo Cyprian Anaene Oyeka, Godday Uwawunkonye Ebuh*

Department of Statistics, Faculty of Physical Sciences, Nnamdi Azikiwe University, Awka, Nigeria
Email: *ablegod007@yahoo.com

ABSTRACT

This paper briefly reviews the Wilcoxon signed rank sum test and proposes a modification. Unlike the Wilcoxon method, the proposed approach does not require that the populations being studied be continuous. Also unlike the Wilcoxon signed rank test the proposed method, does not require the absence of zero differences or tied absolute values of differences. Rather the proposed method structurally makes provisions for these possibilities. The proposed test statistic also enables the estimation of the probabilities of positive, zero or tied and negative differences within the data. This was illustrated with an example and the proposed method was generally more efficient and hence more powerful than the Wilcoxon test statistic with the power increasing as the number of tied observations or zero differences increases.

Keywords: Proposed Method; Probabilities; Positive; Zero; Tied; Negative; Frequencies

1. Introduction

Wilcoxon signed rank test is a rank based alternative to the parametric t test that assumes only that the distribution of differences within pairs be symmetric without requiring normality [1]. Let X_i be the ith observation, $i = 1, 2, \cdots, n$ in a random sample of size n drawn from population X with unknown median M; or let (X_i, Y_i) be the ith pair $i = 1, 2, \cdots, n$ in a paired random sample of size n drawn from population X and Y with unknown M_1 and M_2 respectively. For the moment, we assume that X and Y are continuous. In the one sample case, interest may be in testing that the unknown population median M is equal to some specified value, M_0. In the paired sample case interest may be in testing that the unknown population medians are equal that is $M_1 = M_2$ or that one population median is equal to at least some multiple of other population median, that is $M_1 = c \cdot M_2 + k$ say, where c ($c > 0$) and k are real numbers versus appropriate two-sided or one sided alternative hypotheses. If the assumption of parametric test are satisfied, the first hypothesis may be tested using the one sample t-test while the second hypothesis may be tested using the paired sample t test. The third hypothesis may however be readily tested using the parametric method because of problems of non-homogeneity. If the necessary assumptions of the parametric t-test cannot be reasonably made, use of a non-parametric method that often readily suggests itself in these situations is the Wilcoxon signed rank sum test [2].

In this paper, we briefly discuss the Wilcoxon method and then proceed to present a modified version of the method that may be appropriate for testing the above hypotheses.

2. The Wilcoxon Signed Rank Sum Test

According to [3,4], the Wilcoxon signed rank test is used to test the null hypothesis that the median of a distribution is equal to some value and can be used in place of a one sample t-test, a paired t-test or for ordered categorical data where a numerical scale is inappropriate but where it is possible to rank the observations.

To use the Wilcoxon signed rank sum test, we first find the difference between the observation and the hypothesized median in the one sample problem or the difference between the paired observations in the paired sample problems. That is, in the one sample case, we find $d_i = x_i$ or in the two sample case ($d_i = x_i - cy_i - k$) for $i = 1, 2, \cdots, n$. We then take the absolute values of these differences and rank them either from the smallest to the largest or from the largest to the smallest, always taking note of the ranks of the absolute values with positive differences and those with negative differences. The requirement that the populations from which the samples are drawn are continuous makes it possible to state at least theoretically that the probability of obtaining zero differences or tied absolute values of the differences is zero. Now, let $r\left(|d_i|\right)$ be the rank assigned to $|d_i|$, the absolute value of the ith difference d_i; for $i = 1, 2, \cdots, n$.

Let $Z_i = \begin{cases} 1, & \text{if } d_i > 0; \\ 0, & \text{if } d_i < 0. \end{cases}$ (1)

Let $\theta = P(Z_i = 1)$ (2)

$$T^+ = \sum_{i=1}^{n} Z_i r\left(|d_i|\right) \qquad (3)$$

That is T is the sum of the ranks of the absolute values with positive differences. Now for simplicity but without loss of generality, we let $r\left(|d_i|\right) = i$ [2].

Then,

$$T^+ = \sum_{i=1}^{n} i \cdot Z_i \qquad (4)$$

Now

$$\mathrm{E}(Z_i) = \theta \text{ and } \mathrm{Var}(Z_i) = \theta(1-\theta) \qquad (5)$$

Hence

$$\mathrm{E}\left(T^+\right) = \mathrm{E}\left(\sum_{i=1}^{n} i \cdot Z_i\right) = \sum_{i=1}^{n} i \cdot \mathrm{E}(Z_i)$$

That is

$$\mathrm{E}\left(T^+\right) = \frac{n(n+1)}{2} \cdot \theta \qquad (6)$$

Note that an estimate of θ namely $\hat{\theta}$ may be obtained from the expression

$$T^+ = \frac{n(n+1)}{2} \hat{\theta} \qquad (7)$$

$$\mathrm{Var}\left(T^+\right) = \mathrm{Var} \cdot \left(\sum_{i=1}^{n}(i \cdot Z_i) = \sum_{i=1}^{n}\left(i^2 \cdot \mathrm{Var}(Z_i)\right)\right)$$

Since $\mathrm{Cov}(Z_i, Z_j) = 0$, for $i \neq j$
That is

$$\mathrm{Var}\left(T^+\right) = \frac{n(n+1)(2n+1)}{6} \cdot \theta(1-\theta) \qquad (8)$$

The null hypothesis that is usually tested in the Wilcoxon signed rank sum approach is [2]. $H_0 : \theta = \theta_0 = \frac{1}{2}$ versus either a two sided or an appropriate one sided alternative hypothesis. For the paired sample case, this null hypothesis is equivalent to $H_0 : P = (M_1 < M_2) = \frac{1}{2}$.

Similarly for the one sample case this null hypothesis is equivalent to $H_0 : M = M_0$. A large sample test statistic for any hypothesized value of 0, θ_0 say $(0 < \theta < 1)$ is given by

$$Z = \frac{T^+ - \dfrac{n(n+1)\theta_0}{2}}{\sqrt{\dfrac{n(n+1)(2n+1)}{6}\theta_0(1-\theta_0)}} \qquad (9)$$

This has approximately a standard normal distribution under the desired null hypothesis. But under the null hypothesis usually tested using the Wilcoxon signed rank

test, that is,

$H_0 : \theta = \theta_0 = \dfrac{1}{2}$, then Equation (6) becomes

$$\mathrm{E}\left(\frac{T^+}{H_0}\right) = \frac{n(n+1)}{4} \qquad (10)$$

and $\mathrm{Var}\left(\dfrac{T^+}{H_0}\right) = \dfrac{n(n+1)(2n+1)}{24} \qquad (11)$

Hence the test statistics of Equation (9) becomes

$$Z = \frac{T^+ - \dfrac{n(n+1)}{4}}{\sqrt{\dfrac{n(n+1)(2n+1)}{24}}} \qquad (12)$$

which under H_0 has a standard normal distribution for fairly large sample size n and may be used to test the null hypothesis of equal population medians. When all non zero values are of the same sign, the Wilcoxon signed rank test reduces to the sign test and the two tailed P-value is $\left(\dfrac{1}{2}\right)^{n-1}$, where n is the number of non zero values [5].

3. The Proposed Modified Method

We here drop the requirement that the populations from which the samples are drawn are continuous. We now only require that the populations be quantitative data measured on at most the ordinal scale. The populations could be continuous or discrete. The requirements of no zero differences or tied absolute values are also no longer necessary; since these problems are taken care of structurally by following model specifications. We here continue to use d_i to represent the difference between x_i and the hypothesized population median M_0, in the one-sample case and the differences $x_i - y_i$ (or $x_i - y_i - $ k) in the paired sample case for $i = 1, 2, \cdots, n$. We also continue to use again without loss of generality $r\left(|d_i|\right) = i$ to represent the rank assigned to the absolute value of the ith difference, d_i.

$$Z_i = \begin{cases} 1, & \text{if } d_i > 0; \ 0, \text{ if } d_i = 0; \text{ and} \\ -1, & \text{if } d_i < 0 \text{ for } i = 1, 2, \cdots, n \end{cases} \qquad (13)$$

Also let

$$\pi^+ = P(Z_i = 1), \ \pi^0 = P(Z_i = 0), \ \pi^- = P(Z_i = -1) \qquad (14)$$

where $\pi^+ + \pi^0 + \pi^- = 1 \qquad (15)$

Finally define

$$T = \sum_{i=1}^{n} i \cdot Z_i \qquad (16)$$

That is, T is the difference between the sum of ranks assigned to absolute values with positive differences and the sum of the ranks assigned to absolute values with negative differences.

$$E(Z_i) = (1)\pi^+ + (0)\pi^0 + (-1)\pi^-$$

That is

$$E(Z_i) = (\pi^+ - \pi^-) \qquad (17)$$

and

$$\begin{aligned}\text{Var}(Z_i) &= E(Z_i)^2 - (E(Z_i))^2 \\ &= 1^2\pi^+ + 1^0\pi^0 + (-1)^2\pi^- - (\pi^+ - \pi^-)^2\end{aligned}$$

That is

$$\text{Var}(Z_i) = \pi^+ + \pi^- - (\pi^+ - \pi^-)^2 \qquad (18)$$

Furthermore

$$E(T) = \sum_{i=1}^{n}(i \cdot Z_i) = \sum_{i=1}^{n}(i \cdot E(Z_i))$$

That is

$$E(T) = \frac{n(n+1)}{2} \cdot (\pi^+ + \pi^-) \qquad (19)$$

Note that $\pi^+ + \pi^-$ may be estimated as $\hat{\pi}^+ + \hat{\pi}^-$ from

$$T = \frac{n(n+1)}{2(\hat{\pi}^+ + \hat{\pi}^-)} \qquad (20)$$

Also $\text{Var}(T) = \text{Var} \cdot \left(\sum_{i=1}^{n}(i \cdot Z_i) = \sum_{i=1}^{n}(i^2 \cdot \text{Var}(Z_i)) \right)$

Since $\text{Cov}(Z_i, Z_j) = 0$, for $i \neq j$
Therefore

$$\text{Var}(T) = \frac{n(n+1)(2n+1)}{6} \cdot (\pi^+ + \pi^- - (\pi^+ - \pi^-)^2) \qquad (21)$$

Note that in the case of one sample π^+, π^0 and π^- are respectively on the average the probabilities that the population median is greater than, equal to, or less than the hypothesized median Mo while in the paired sample case, they are on the average the probabilities that one population median is greater than, equal to or less than the other population median. They are estimated respectively as the relative frequencies of occurrence of 1, 0 and -1 in the frequency distribution of the n elements of Z_i.

That is $\hat{\pi}^+ = \dfrac{f^+}{n}$; $\hat{\pi}^0 = \dfrac{f^0}{n}$; $\hat{\pi}^- = \dfrac{f^-}{n}$ (22)

where f^+, f^0 and f^- are respectively the frequencies of occurrence of 1, 0 and -1 in the frequency distri-

bution of Z_i. Often the null hypothesis required to be tested is

$$H_0 : \pi^+ - \pi^- = \sigma_0 \text{ versus } H_1 : \pi^+ - \pi^- > \sigma_0,$$
$$\text{say, } (-1 < \sigma_0 < 1) \qquad (23)$$

or versus a two tailed or an appropriate one tailed alternative hypothesis. For the paired sample case, this null hypothesis is equivalent to

$$H_0 : P(M_1 > M_2) - P(M_1 < M_2) = \sigma_0$$

which for $\sigma_0 = 0$ is the same as

$$H_0 : M_1 = M_2 \text{ or } (M_1 = cM_2 + k)$$

For the one sample case, the null hypothesis is equivalent to

$$H_0 : P(M > M_0) - P(M < M_0) = \sigma_0$$

which for $\sigma_0 = 0$ implies $H_0 : M = M_0$

The test statistic

$$Z = \frac{T - \dfrac{n(n+1)}{2}\sigma_0}{\sqrt{\dfrac{n(n+1)(2n+1)}{6}(\pi^+ + \pi^- - (\pi^+ - \pi^-)^2)}} \qquad (24)$$

which under H_0 has a standard normal distribution for fairly large n. In practical application π^+ and π^- of Equation (24) are usually replaced with their sample estimates of Equation (22). Note that T^+ as defined by Wilcoxon is the sum of the ranks of the absolute values of positive differences and that θ is the probability of the occurrence of only positive differences. Thus the specification of Wilcoxon's statistic does not explicitly provide for the possible occurrence of negative differences. So the π^- in our proposed modification in Equation (14) is automatically set equal to zero in the Wilcoxon's approach and under H_0 the probability of the occurrence of positive differences is hypothesized as

$H_0 : \theta = \theta_0 = \dfrac{1}{2}$. Hence if we automatically set

$\pi^- = 0$ and under the null hypothesis we set

$\pi^+ = \pi_0^+ (= \sigma_0) = \theta_0 = \dfrac{1}{2}$ in Equation (24), then the proposed test statistic is seen to coincide with Wilcoxon signed rank test statistic could be rewritten as

$$Z = \frac{4T - n(n+1)}{\sqrt{\dfrac{2n(n+1)(2n+1)}{3}(\pi^+ + \pi^- - (\pi^+ - \pi^-)^2)}} \qquad (25)$$

while the test statistic for the proposed method could be rewritten as

$$Z = \frac{2T - n(n+1)\sigma_0}{\sqrt{\dfrac{2n(n+1)(2n+1)}{3}(\pi^+ + \pi^- - (\pi^+ - \pi^-)^2)}} \qquad (26)$$

The hypothesis tested under the Wilcoxon's approach $H_0 : \theta = \theta_0 = \dfrac{1}{2}$ is equivalent to $H_0 : \pi^+ - \pi^- = \sigma_0 = 0$ tested under the proposed method.

Therefore the relative rate at which these two test statistics are able to accept a true null hypothesis or reject a false null hypothesis can therefore be obtained by comparing the variance of $4T^+$ of Equation (25) with the variance of $2T$ of Equation (26) that is in terms of the relative efficiency of T compared with T^+ that is

$$\text{RE}\left(T;T^+\right) = \frac{\text{Var}\left(4T^+\right)}{\text{Var}\left(2T\right)} = \frac{4\text{Var}\left(T^+\right)}{2\text{Var}\left(T\right)}$$
$$= \frac{1}{\pi^+ + \pi^- - \left(\pi^+ - \pi^-\right)^2} \qquad (27)$$

In other words

$$\text{RE}\left(T;T^+\right) = \frac{1}{\pi^+ + \pi^- - \left(\pi^+ - \pi^-\right)^2} \geq \frac{1}{\left(\pi^+ + \pi^-\right)}$$

since $\left(\pi^+ - \pi^-\right)^2 \geq 0$

That is

$$\text{RE}\left(T;T^+\right) \geq \frac{1}{\left(1 - \pi^0\right)} \qquad (28)$$

Since from Equation (15), $\pi^+ + \pi^- = 1 - \pi^0$

Therefore

$$\text{RE}\left(T;T^+\right) \geq 1 \qquad (29)$$

for $\pi^0 \geq 0$, increasing as π^0 increases.

Hence, the proposed modified (T) is more efficient than the Wilcoxon's method $\left(T^+\right)$ whenever π^0 is not equal to 0, that is whenever there are zero differences or tied observations in the data. The relative efficiency of T compared with T^+ increases as π^0 increases. The two methods are equally efficient only when $\pi^0 = 0$, that is when there are no ties in the data whatsoever.

Thus unless there are no ties whatsoever in the data, the proposed method yields a more powerful test criterion than the Wilcoxon signed rank sum test statistic for the same sample size.

4. Illustrative Example

A random sample of twelve married women were selected and asked to state the actual and the ideal number of children they would like to have. The results are presented in **Table 1**.

To apply the Wilcoxon signed rank test, we take and rank the differences between the actual and ideal number of children by the sample of married women shown in **Table 1**. Tied absolute differences are assigned their mean ranks. The results are presented in **Table 2**.

Table 1. Actual and ideal number of children by a sample of married women.

Woman	Actual	Ideal
1	4	5
2	1	5
3	6	5
4	1	6
5	7	5
6	1	9
7	4	4
8	2	6
9	8	8
10	5	5
11	4	4
12	4	5

Table 2. Ranks of absolute differences $r\left(|d_i|\right)$, of the differences d_i between actual and ideal number of children in Table 1.

No. of Women	1	2	3	4	5	6	7	8	9	10	11	12
Ideal	5	5	5	6	5	9	4	6	8	5	4	5
Actual	4	1	6	1	7	1	4	2	8	5	4	4
d_i (Ideal-Actual)	1	4	–1	5	–2	8	0	4	0	0	0	1
$|d_i|$	1	4	1	5	2	8	0	4	0	0	0	1
Rank of $|d_i|$	6	9.5	6	11	8	12	2.5	9.5	2.5	2.5	2.5	6
Sign of $|d_i|$	+	+	–	+	–	+	0	+	0	0	0	+

From **Table 2** we have that the sum of the ranks of absolute differences with positive signs is

$$T^+ = 6 + 9.5 + 11 + 12 + 9.5 + 6 = 54$$

The null hypothesis to be tested with the Wilcoxon signed rank test is usually that the two populations of interest have equal medians $\left(H_0 : \theta = \theta_0 = 0.5\right)$. Hence under $\left(H_0\right)$ we obtain from Equation (6) and (8) that

$$\text{E}\left(T^+\right) = \frac{12(13)}{4} = 39 \quad \text{and}$$

$$\text{Var}\left(T^+\right) = \frac{12(13)(25)}{24} = 162.5 \quad \left(H_0 : \theta = \theta_0 = 0.5\right)$$

The resulting test statistic (Equation (12)) under the Wilcoxon approach is

$$Z = \frac{54 - 39}{\sqrt{162.50}} = \frac{15}{(12.748)} = 1.177 = 1.18$$
$$\text{(P-value} = 0.1190)$$

which is not statistically significant at the 5 percent level.

We now apply the modified Wilcoxon signed rank test to the data of **Table 1** for comparative purposes. Now from **Table 2** and Equation (16) we have that

$T = T^+ - T^- = 54 - (6 + 8) = 40$ and from Equation (22)

$$\hat{\pi}^+ = \frac{6}{12} = 0.50 \;;\;\; \hat{\pi}^0 = \frac{4}{12} = 0.33 \;;\;\; \hat{\pi}^- = \frac{2}{12} = 0.17$$

from Equations (19) and (21), we have that

$$E(T) = \frac{12(13)}{2} \times (0.50 - 0.17) = 25.74$$

and

$$\text{Var}(T) = \frac{12(13)(25)}{6} \times \left(0.50 + 0.17 - (0.50 - 0.17)^2\right)$$
$$= 364.715$$

To test the null hypothesis of equal population medians using the modified approach we have from Equation (24), with $\sigma_0 = 0$

$$Z = \frac{40}{\sqrt{364.715}} = \frac{40}{(19.098)} = 2.094$$

(P-value = 0.0183)

which is statistically significant at 5 percent level.

Hence we now reject the null hypothesis that the two populations have equal medians. Note that this same null hypothesis was accepted using the Wilcoxon signed rank test statistic unmodified for possible presence of ties in the data.

Also note from Equation (28) that the relative efficiency of the modified test statistic T to the Wilcoxon test statistic T^+ is estimated as

$$\text{RE}(T; T^+) \geq \frac{1}{(1 - \pi^0)} = \frac{1}{(1 - 0.33)} = \frac{1}{0.67} = 1.49$$

Thus for the data being analysed, the proposed test statistic is at least 1.49 times more efficient and hence more powerful than the Wilcoxon signed rank test statistic T^+.

5. Conclusion

This paper briefly reviews the Wilcoxon signed rank sum test and proposes a modification. Unlike the Wilcoxon method, the proposed approach does not require that the populations being studied be continuous. Also unlike the Wilcoxon signed rank test the proposed method, does not require the absence of zero differences or tied absolute values of differences. Rather the proposed method structurally makes provisions for these possibilities. The proposed test statistic also enables the estimation of the probabilities of positive, zero or tied and negative differences within the data. The proposed method shown to be generally more efficient and hence more powerful than the Wilcoxon test statistic with the power increasing as the number of tied observations or zero differences increases.

REFERENCES

[1] R. De-W. Derryberry, S. B. Schou and W. J. Conover, "Teaching Rank-Based Tests by Emphasizing Structural Similarities to Corresponding Parametric Tests," *Journal of Statistics Education*, Vol. 18, No. 1, 2010, pp. 1-19. www.amstat.org/publications/jse/v18n1/derryberry.pdf

[2] D. Gibbon, "Non Parametric Statistics," McGraw Hill, New York, 1971. onlinelibrary.wiley.com/doi/10.1111/j.2044-8317

[3] R. Shier, "Statistics," The Wilcoxon Signed Rank Sum Test, Mathematics Learning Support Centre, 2004, pp. 1-3. mlsc.lboro.ac.uk/resources/statistics/signtest.pdf

[4] I. C. A. Oyeka, "An Introduction to Applied Statistical Methods," 8th Edition, Nobern Avocation Publishing Company, Enugu, 2009, pp. 496-533.

[5] B. H. Robbins, "Non Parametric Tests," Scholars Series, 2010, pp.1-30. biostat.mc.vanderbilt.edu/wiki/pub/Main/.../NonParametrics.pdf