Clinical Medicine and Health Research Journal (CMHRJ)
Volume 04, Issue 06, November - December, 2024
Page No. 1037-1045,

Research Article

Intrinsically Ties Adjusted Median Test for Determining the Lengths of Hospitalization of Sampled Patients with Hypertension and Malaria in A Population

Okeh, Uchechukwu Marius[1,2] and Emeji Emmanuel Ogwah[3]

[1]Senior Lecturer, Department of Industrial Mathematics and Health Statistics, David Umahi Federal University of Health Sciences Uburu Ebonyi State Nigeria.

[2]International Institute for Nuclear Medicine and Allied Health Research, David Umahi Federal University of Health Sciences Uburu Ebonyi State, Nigeria.

[3]Principal Lecturer, Department of Integrated Science, Ebonyi State College of Education Ikwo in Affiliation with Ebonyi State University Abakaliki Nigeria.

Received: 10 October, 2024 : Accepted: 14 November, 2024 : Published: 20 November 2024

Abstract:

Background: This paper proposes and presents a nonparametric statistical method for the analysis of two sample data that intrinsically and structurally adjusts the test statistic, for the possible presence of tied observations in the sample populations.

Methodology: The proposed procedure makes it unnecessary to require the populations to be continuous as is often the case with some other methods. The populations may be measurement on as low as the ordinal scale and need not be continuous or even numeric. In situations where the original or initial null hypothesis is rejected, test statistics are developed to help determine which of the two populations of interest may be responsible for the rejection of the null hypothesis, an approach that is not possible for some other existing two sample median test.

Results: The proposed method is illustrated with some sample data and shown using the data to compare favorably with the usual median test and the Mann-Whitney U-test that could be used for the same purpose. Result showed that we rejected the null hypothesis, H_0 that hypertension and malaria patients from the population admitted to a hospital for treatment do not have equal median lengths of hospitalization for hypertension and malaria. Since $\chi^2 = 217 > \chi^2_{0.95;1} = 3.841.$ The Chi-square value for testing the null hypothesis H_0 that the median length of hospitalization of hypertension patients in the population is equal to the median length of hospitalization of both hypertension and malaria patients in the population when pooled together as one population, is $\chi^2 = 2.195(P-value = 4.345)$ which with 1 degree of freedom is not statistically significant at the 5 percent significance level $(\chi^2_{0.95;1} = 3.841)$ leading to the non-rejection of the null hypothesis. the Chi-Square value for testing the same null hypothesis with respect to malaria patients, that is that the median length of hospitalization of malaria patients in the population is the same as the median length of hospitalization of the combined or pooled population of hypertension and malaria patients when combined and treated as one population is $\chi^2 = 9.655(P-value = 0.2134)$ which with 1 degree of freedom is statistically significant at the 5 percent significance level $(\chi^2_{0.95;1} = 3.841)$ leading to the rejection of the null hypothesis.

Conclusions and recommendations: We conclude that hypertension and malaria patients in the population do not have equal median lengths of hospitalization. We may therefore conclude that the median length of hospitalization of malaria patients is statistically different from the median length of hospitalization of both hypertension and malaria patients in the sampled population and may hence be responsible for the rejection of the initial null hypothesis H0 of Equation 11 or 12 of equal population median lengths of hospitalization of the two types of patients in the sampled population. since the Chi-square value of $\chi^2 = 217.00$ obtained using the proposed modified ties adjusted median test for two samples is much larger than the Chi-square value of $\chi^2 = 8.016$ obtained using the usual ordinary unmodified ties unadjusted two sample median test, the proposed method is likely to correctly reject a false null hypothesis more often and hence is more powerful than the ordinary median test when used to analyze the same sample observations.

Keywords: Intrinsically, Structurally, Mann-Whitney U-test, tied observations, two sample median test.

Introduction

A problem with the usual median test is that it requires the sampled populations to be continuous. This is to ensure at least theoretically that the probability that any two observations have equal values or that any observation is equal to the common median of the sampled populations is zero (Gibbons,1993; Gibbon and Chakraborti,2003). However, practice tied observations do occur between the sampled populations and some observations may be equal to the common population median. Too many of these tied observations if not structurally adjusted for may seriously compromise the power of the test statistic and lead to erroneous conclusions (Spegel,1988; Spegel and Castellan,1988). Furthermore the usual median test for two samples does not immediately enable the researcher, in cases where the initial null hypothesis is rejected, to determine which of the sampled populations may have led to its rejection(Oyeka et al,2009;2010;2011).In this paper we propose and develop a modification of the usual median test for two samples that intrinsically and structurally adjusts the test statistic for the possible presence of tied observations between the sampled populations and for situations in which some observations are equal to the common median of the two populations. This approach obviates the need to require the sampled populations to be continuous. The populations may now be measurements on as low as the ordinal scale and need not be continuous or even numeric.

Unlike its currently existing counterpart, would also easily enable the researcher determine in cases in which a null hypothesis of interest is rejected which of the two sampled populations is likely to have led to rejection of the initial null hypothesis.

Materials and Method

Let x_{hj} be the hth observation or score in a random sample of size hj independently drawn from population j for $h = 1, 2, ..., hj; j = 1, 2.$ Populations x_1 *and* x_2 may be measurements on as low as the ordinal scale and need not be continuous or even numeri. Now to develop the proposed method we may let

$$u_{hj} = \begin{cases} 1. if\ x_{h} > x_{j_2}\ or\ if\ x_{h}\ is\ greater(larger, more, higher, better)\ than\ x_{j_2} \\ 0. \quad if\ x_{h} = x_{j_2}\ or\ if\ x_{h}\ is\ same\ as(equal\ to)\ x_{j_2} \\ -1. \quad if\ x_{h} < x_{j_2}\ or\ if\ x_{h}\ is\ lower(smaller, less, worse)\ than\ x_{j_2} \end{cases} \qquad (1)$$

$$for\ h = 1, 2, ..., n_1; j = 1, 2, ..., n_2$$

Let

$$\pi^{+} = P(u_{hj} = 1); \pi^{0} = P(u_{hj} = 0); \pi^{-} = P(u_{hj} = -1) \qquad (2)$$

Where

$$\pi^{+} + \pi^{0} + \pi^{-} = 1 \qquad (3)$$

Note that the specifications in equations 1-3 have intrinsically and structurally provided adjustments for the possible presence of tied observations between the sampled populations thereby making it unnecessary to require these populations to be continuous or even numeric measurements (Oyeka et al,2009; Oyeka,2013; Ebuh and Oyeka, 2012; Cordel and Foreman,2004).

Now define

$$W = \sum_{h=1}^{n_1} \sum_{j=1}^{n_2} u_{hj} \qquad (4)$$

The expected value or mean and variance of u_{hj} are respectively

$$E\left(u_{hj}\right) = \pi^{+} - \pi^{-}; Var\left(u_{hj}\right) = \pi^{+} + \pi^{-} - \left(\pi^{+} - \pi^{-}\right)^{2} \qquad (5)$$

Also the expected value of W is

$$E\left(W\right) = \sum_{h=1}^{n_1} \sum_{j=1}^{n_2} E\left(u_{hj}\right) = n_1.n_2\left(\pi^{+} - \pi^{-}\right) \qquad (6)$$

And the variance of W is

$$Var\left(W\right) = \sum_{h=1}^{n_1} \sum_{j=1}^{n_2} Var\left(u_{hj}\right) = n_1.n_2\left(\pi^{+} + \pi^{-} - \left(\pi^{+} - \pi^{-}\right)^{2}\right) = n_1.n_2\left(\pi^{+} + \pi^{-}\right) - \frac{W^{2}}{n_1.n_2} \qquad (7)$$

Now π^{+}, π^{0} *and* π^{-} are respectively the probabilities that a randomly selected observation or score from population x_1 is greater(larger, higher, more, better)equal to (the same as)or lower(smaller, less, worse)than a randomly selected observation or score from population x_2.

Their sample estimates are respectively

$$\hat{\pi}^{+} = \frac{F^{+}}{n_1.n_2}; \hat{\pi}^{0} = \frac{F^{0}}{n_1.n_2}; \hat{\pi}^{-} = \frac{F^{-}}{n_1.n_2} \qquad (8)$$

Where F^{+}, F^{0} *and* F^{-} are respectively the number of times in which sample observation drawn from population x_1 are greater(larger, higher, more, better)the same as(equal to)or smaller (lower, less)than sample observations drawn from population x_2. In other words, F^{+}, F^{0} *and* F^{-} are respectively the number of 1's,0's and -1's in the frequency distribution of the $n_1.n_2$ values of these number in u_{hj},

$$For\ h = 1, 2, ..., n_1; j = 1, 2, ..., n_2.$$

Note that $\pi^{+} - \pi^{-}$ measures by how much on the average observations or scores in population x_1 are greater(larger, higher, more, better) less the probability that is lower(smaller, less, worse)than observations or scores in population x_2 or equivalently, $\pi^{+} - \pi^{-}$ is a measure of the proportion by which or the probability that the score of a randomly selected subject from population x_1 is higher (greater, more, larger, better)less the probability that the randomly selected subject's score is smaller(lower, less, worse)than the score by a randomly selected subject from population x_2. Its sample estimate is

$$\hat{\pi}^+ - \hat{\pi}^- = \frac{W}{n_1 n_2} = \frac{F^+ - F^-}{n_1 n_2} \quad (9)$$

The estimated variance of $\pi^+ - \pi^-$ is from Equations 5-7

$$Var\left(\hat{\pi}^+ - \hat{\pi}^-\right) = \frac{Var(W)}{n_1^2 n_2^2} = \frac{\hat{\pi}^+ + \hat{\pi}^- - \left(\hat{\pi}^+ - \hat{\pi}^-\right)^2}{n_1 n_2} \quad (10)$$

Now the general null hypothesis to be tested here is that the population medians m_{10} *and* m_{20} of populations x_1 *and* x_2 respectively differ by some value m_{d0} is equivalent to the null hypothesis H_0.

$$H_0: \pi^+ - \pi^- = \theta_0 \ versus \ H_1: \pi^+ - \pi^- \neq \theta_0 \left(-1 \le \theta_0 \le 1\right) \quad (11)$$

Note that if the population medians m_{10} *and* m_{20} of populations x_1 *and* x_2 are equal to each other then θ_0 would be zero so that the two populations would be expected to have equal population medians m_0. Thus if $\theta_0 = 0$, then the null Otherwise the null hypothesis H0 is rejected is accepted. The proposed test statistic because of the adjustment intrinsically made for the possible presence of ties in the data is more efficient and hence likely to be more powerful than the ordinary median test unadjusted for ties between the population. The efficiently and power of the test increases as the number of tied observations and hence as the probability of tied observations and hence as the probability of ties π_0 increases (Oyeka et al, 2009; Friedlin and Gastwirth,2000). There is also another interesting aspect of the proposed modified ties adjusted median test for two sample that is not possible with the ordinary median test that is unadjusted for ties in the data when the null hypothesis H0 of equal population medians is rejected, the modified ties adjusted median test unlike the ordinary median test can be further used to determine which of the two populations may have led to the rejection of the initial null hypothesis H_0 of equations 11 or 12 (Oyeka et al,2010; Oyeka,2013).The rationale behind this approach is that if in fact the two populations have equal medians m_{10} *and* m_{20},then each of these population medians would be expected to be equal to the common population median m_0 of the population which would have been obtained if the two populations were pooled into one combined population. In other words, if the two populations x_1 *and* x_2 have equal population medians, then each of these medians is expected to be equal to the common medians m_0 of the population from which the combined or pooled samples in the median test would actually have been randomly drawn. Thus, suppose m_{10}, m_{20} *and* m_0 are respectively the often, unknown population medians of populations x_1 *and* x_2 and of the population obtained by pooling or combining these two populations into one combined

hypothesis H_0 of Equation 11 is equivalent to the null hypothesis

$$H_0: m_{10} = m_{20} = m_0 \ versus \ H_1: m_{L0} \neq m_0 \quad (12)$$

For some L=1,2,where m_0 is the unknown common population median of populations x_1 *and* x_2 .The null hypothesis H0 of equations 11 or 12 is tested using the test statistic

$$\chi^2 = \frac{\left(w - n_1 n_2 \theta_0\right)^2}{Var(w)} = \frac{n_1 n_2\left[\left(\hat{\pi}^+ - \hat{\pi}^-\right) - \theta_0\right]^2}{\hat{\pi}^+ + \hat{\pi}^- - \left(\hat{\pi}^+ - \hat{\pi}^-\right)^2} \quad (13)$$

Which under H0 and for sufficiently large sample sizes $n_1, n_2 \left(n_1 \ge 8; n_2 \ge 8\right)$ (Gibbons,1993) has approximately the Chi-square distribution with 1 degree of freedom. The null hypothesis H0 of equations 11 or 12 is rejected at the α level of significance if

$$\chi^2 \ge \chi_{1-\alpha;1} \quad (14)$$

population. Now if in fact any two populations x_1 *and* x_2 have equal population medians m_0 ,then approximately one half(in the absence of ties) of the observations in the random samples drawn from each of these populations x_1 *and* x_2 would be expected to lie above the common median m of these samples when combined into are sample appropriately one half of the observations from each of the samples would be expected to lie below the common sample median m. In general if the populations have equal population, median m_0 then approximately equal proportions of the observation of random samples drawn from each of these populations would be expected to lie above or below the common median m of the combined sample drawn from the populations. In particular x_1 with median m_{10} and population x_2 with median m_{20} have equal medians m_0 ,then approximately equal proportions of observations of random samples drawn from each of them with sample medians m_1 *and* m_2 respectively with common combined or pooled sample median m would be expected to lie above or below the common sample median m of the pooled sample therefore to ascertain whether for instance population x_1 has the same median m_{10} as the combined population with common population median m_0 we may let

$$u_{i1} = \begin{cases} 1, if\ x_{i1} > m\ or\ if\ x_{i1}\ is\ greater\ (larger, higher, more, better)\ score\ than\ m \\ 0, \quad if\ x_{i1} = m\ or\ if\ x_{i1}\ is\ the\ same\ as\ (equal\ to)\ score\ than\ m \\ -1, \quad if\ x_{i1} < m\ or\ if\ x_{i1}\ is\ lower\ (smaller, less, worse)\ score\ than\ m \end{cases} \quad (15)$$

Where x_{i1} is the ith observation in a random sample of size n_1 drawn from population x_1 for i=1,2,…, n_1.

Let

$$\pi^+_{x_1} = P(u_{ix_1} = 1); \pi^0_{x_1} = P(u_{ix_1} = 0); \pi^-_{x_1} = P(u_{ix_1} = -1) \quad (16)$$

Where

$$\pi^+_{x_1} + \pi^0_{x_1} + \pi^-_{x_1} = 1 \quad (17)$$

Also let

$$w_{x_1} = \sum_{i=1}^{n_1} u_{ix_1} \quad (18)$$

Now the expected value or mean an variance of u_{ix_1} are respectively

$$E(u_{ix_1}) = \pi^+_{x_1} - \pi^-_{x_1}; Var(u_{ix_1}) = \pi^+_{x_1} + \pi^-_{x_1} - (\pi^+_{x_1} - \pi^-_{x_1})^2 \quad (19)$$

Also the expected value of w_{x_1} is

$$E(w_{x_1}) = n_1(\pi^+_{x_1} - \pi^-_{x_1}) \quad (20)$$

And the variance of w_{x_1} is

$$Var(w_{x_1}) = \sum_{i=1}^{n_1} Var(u_{ix_1}) = n_1\left(\pi^+_{x_1} + \pi^-_{x_1} - (\pi^+_{x_1} - \pi^-_{x_1})^2\right) \quad (21)$$

Now $\pi^+_{x_1}, \pi^0_{x_1}$ *and* $\pi^-_{x_1}$ are respectively the proportions or the probabilities that a randomly selected observations or score from population x_1 is greater (larger, more, higher, better)equal to (the same as) or smaller(lower, less, worse)than the common sample median m of the combined or pooled sample observations from populations x_1 *and* x_2.Their sample estimates are respectively

$$\hat{\pi}^+_{x_1} = \frac{F^+_{x_1}}{n_1}; \hat{\pi}^0_{x_1} = \frac{F^0_{x_1}}{n_1}; \hat{\pi}^-_{x_1} = \frac{F^-_{x_1}}{n_1} \quad (22)$$

Where $F^+_{x_1}, F^0_{x_1}$ *and* $F^-_{x_1}$ are respectively the number of times sample observations or scores from population x_1 are greater(larger, more, higher, better)equal to (the same value as) or smaller (lower, less, worse)than m, the sample median of the combined or pooled sample observation from populations x_1 *and* x_2.

In other words $F^+_{x_1}, F^0_{x_1}$ *and* $F^-_{x_1}$ are respectively the number of 1's,0's and -1's in the frequency distribution of the n_1 values of these numbers in u_{ix_1} for i=1,2,…, n_1.

Where,

$$F^+_{x_1} + F^0_{x_1} + F^-_{x_1} = n_1 \quad (23)$$

Note that $\pi^+_{x_1} - \pi^-_{x_1}$ is estimated as

$$\hat{\pi}^+_{x_1} - \hat{\pi}^-_{x_1} = \frac{w_{x_1}}{n_1} = \frac{F^+_{x_1} - F^-_{x_1}}{n_1} \quad (24)$$

Whose estimated variance is from equation 21

$$Var(\hat{\pi}^+_{x_1} - \hat{\pi}^-_{x_1}) = \frac{Var(w_{x_1})}{n_1^2} = \frac{\hat{\pi}^+_{x_1} + \hat{\pi}^-_{x_1} - (\hat{\pi}^+_{x_1} - \hat{\pi}^-_{x_1})^2}{n_1} \quad (25)$$

Now as noted above if the two samples are in fact drawn from populations with equal population medians m_0 then we would expect at least theoretically(in the absence of ties) that approximately one half of the observations in each sample will lie above the common median m and one half will lie below m, the common sample median of the two samples drawn from populations x_1 *and* x_2 when pooled together as one combined sample; that is in practice we would expect that approximately equal proportions of each sample will lie above as below m. In particular and for the problem at hand with adjustments made for tied observations equal proportions of the random sample drawn from population x_1 and of the random sample drawn from population x_2 are expected to lie above as below the common median m.

Equivalently if the random sample drawn from population x_1 with unknown population median, m_{10} has the same median m with the pooled random sample from populations x_1 *and* x_2, then one would expected that the unknown median $m_c = m_0$ of the population from which the pooled random sample could have been drawn would be equal to m_{10}, this suggest that the null hypothesis to be tested for the random sample drawn from population x_1 here is

$$H_0 : \pi^+_{x_1} = \pi^-_{x_1} \text{ or } H_0 : \pi^+_{x_1} - \pi^-_{x_1} = 0 \quad versus$$

$$H_1 : \pi^+_{x_1} - \pi^-_{x_1} \neq 0 \quad (26)$$

This is the same as testing the null hypothesis $H_0 : m_{10} = m_c = m_0$ versus an appropriate two sided or one sided alternative hypothesis H1.Under the null hypothesis of equation 26,the test statistic

$$\chi^2 = \frac{(\hat{\pi}^+_{x_1} - \hat{\pi}^-_{x_1})^2}{Var(\hat{\pi}^+_{x_1} - \hat{\pi}^-_{x_1})} = \frac{w^2_{x_1}}{Var(w_{x_1})} = \frac{n_1(\hat{\pi}^+_{x_1} - \hat{\pi}^-_{x_1})^2}{\hat{\pi}^+_{x_1} + \hat{\pi}^-_{x_1} - (\hat{\pi}^+_{x_1} - \hat{\pi}^-_{x_1})^2} \quad (27)$$

under the null hypothesis H0 has approximately the Chi-square distribution with 1 degree of freedom for sufficiently large sample

size $n_1 (n_1 \geq 8)$.The null hypothesis H0 of equation 26 is

rejected at the α level of significance if equation 14 is satisfied, otherwise
H0 is accepted in which case one would be able to conclude that the population median m_{10} of population x_1 is not different from the population median m_0 of the combined or pooled observation in populations x_1 *and* x_2. A rejection of the null hypothesis H0 of equation 26 at a chosen α level of significance would suggest that population x_1 and the population that would result as a consequence of pooling populations x_1 *and* x_2 do not have equal population medians, leading to the conclusion that population x_1 is probably responsible for the rejection of initial null hypothesis H0 of equations 11 or 12 that populations x_1 *and* x_2 have equal population medians. Equivalent expressions for population x_2 may be similarly obtained by simply replacing 1 with 2 an n_1 *with* n_2 in equations 15 to 27.Further analysis may now continue without encountering any new problems in the process. The required null hypothesis H0 for population x_2 is then

$$H_0: \pi^{+}_{x_2} = \pi^{-}_{x_2} \; or \; H_0: \pi^{+}_{x_2} - \pi^{-}_{x_2} = 0 \; versus \; H_1: \pi^{+}_{x_2} - \pi^{-}_{x_2} \neq 0 \qquad (28)$$

This is the same as testing the null hypothesis

$$H_0: m_{20} = m_c = m_0 \; versus \; H_1: m_{20} \neq m_c = m_0, say \qquad (29)$$

The null hypothesis H0 of Equations 28 or 29 is tested using the test statistic

$$\chi^2 = \frac{\left(\hat{\pi}^{+}_{x_2} - \hat{\pi}^{-}_{x_2}\right)^2}{Var\left(\hat{\pi}^{+}_{x_2} - \hat{\pi}^{-}_{x_2}\right)} = \frac{w^2_{x_2}}{Var\left(w_{x_2}\right)} = \frac{n_2\left(\hat{\pi}^{+}_{x_2} - \hat{\pi}^{-}_{x_2}\right)^2}{\hat{\pi}^{+}_{x_2} + \hat{\pi}^{-}_{x_2} - \left(\hat{\pi}^{+}_{x_2} - \hat{\pi}^{-}_{x_2}\right)^2} \qquad (30)$$

Which under the null hypothesis H0 of Equations 28 or 29 has approximately the Chi-square distribution with 1 degree of freedom for sufficiently large sample size $n_2 (n_2 \geq 8)$. The null hypothesis H0 of Equations 28 or 29 is rejected at the α level of significance if Equation 14 is satisfied, otherwise H0 is accepted, in which case one would be able to conclude that the population median m_{20} of population x_2 is probably equal to the common population median $m_c = m_0$ of the combined or pooled observation. In populations x_1 *and* x_2, a rejection of the null hypothesis H0 of Equations 28 or 29 would suggest that population x_2 and the population that would result as a consequence of combining or pooling together observations from populations x_1 *and* x_2 are likely to have unequal population medians leading to the conclusion that population x_2 is probably responsible for the rejection of the initial null hypothesis H0 of Equation 11 or 12 of equal population medians. However if the null hypothesis H0 of Equations 26 and 28 or 29 are both accepted we would then be able to conclude that populations x_1 *and* x_2 have equal population medians m_0 otherwise we would conclude that the two population medians m_{10} *and* m_{20} differ from the common population median m_0. If on the other hand only one of the null hypothesis H0 is rejected we would then be able to conclude that the concerned population has a median different from the common population median $m_c = m_0$ and hence may be responsible for the rejection of the overall null hypothesis H0 of equality of the two populations medians. As noted earlier, unlike is the case with the proposed modified ties adjusted two sample median test, it is not possible to use the ordinary median test that is unadjusted for ties for these additional analysis following a rejection of the initial null hypothesis of equal population medians, which is a useful and added advantage of the proposed method over and above the usual ties unadjusted two sample median test.

Results

The following data are the lengths of hospitalization (in days)of random samples of patients admitted for two types of illnesses(Hypertension, x_1 and Malaria, x_2)in a certain population(Table 1)

Table 1: Lengths of hospitalization in days of patients admitted for two type of illnesses in a population.

S/N	Malaria	Hypertension	Pooled sample x_{i2} *and* x_{i1}		Rank of observation in pooled sample u_{ij}	
	x_{i2}	x_{i1}			R_{i2}	R_{i1}
1	7	7	7	18. 7	20.5(m=7days)	20.5
2	11	4	11	19. 4	29	10
3	2	9	2	20. 9	2.5	25.5
4	7	18	7	21. 18	20.5	37
5	3	5	3	22. 5	6	14
6	7	17	7	23. 17	20.5	35.5
7	3	17	3	24. 17	6	35.5
8	1	6	1	25. 6	1	17
9	7	5	7	26. 5	20.5	14
10	8	13	8	27. 13	24	32.5
11	5	16	5	28. 16	14	34
12	3	10	3	29. 10	6	27.5
13	3	12	3	30. 12	6	30.5
14	2	7	2	31. 7	2.5	20.5
15	5	5	5	32. 5	14	14
16	3	12	3	33. 12	6	30.5
17	4	13	4	34. 13	10	25.5
18		9		35. 9		10
19		4		36. 4		27.5
20		10		37. 10	**209(** R_2 **)**	**494(** R_1 **)**

We use the sample data of Table 1 to illustrate the proposed method namely modified ties adjusted two sample median test. Now using the specification of Equation 1, we obtain the values of u_{hj} shown in Table (2).

Table 2: Coded comparison of lengths of hospitalization by Type of illness u_{hj} (Equation 1)

Malaria Patients x_{i2} (in days)

Hypertension Patient x_{i1} (in days)	7	11	2	7	3	7	3	1	7	8	5	3	3	2	5	3	4
7	0	-1	1	0	1	0	1	1	0	-1	1	1	1	1	1	1	1
4	-1	-1	1	-1	1	-1	1	1	-1	-1	-1	1	1	1	-1	1	0
9	1	-1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
18	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5	-1	-1	1	-1	1	-1	1	1	-1	-1	0	1	1	1	0	1	1
17	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
17	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6	-1	-1	1	-1	1	-1	1	1	-1	-1	1	1	1	1	1	1	1
5	-1	-1	1	-1	1	-1	1	1	-1	-1	0	1	1	1	0	1	1
13	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
10	1	-1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
7	0	-1	1	0	1	0	1	1	0	-1	1	1	1	1	1	1	1
5	-1	-1	1	-1	1	-1	1	1	-1	-1	0	1	1	1	0	1	1
12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
13	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9	1	-1	1	1	1	-1	1	1	1	1	1	1	1	1	1	1	1
4	-1	-1	1	-1	1	-1	1	1	-1	-1	-1	1	1	1	1	1	0
10	1	-1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

From Table 2 we have that $F^+ = 274; F^0 = 16; F^- = 50$

So that $w = F^+ - F^- = 274 - 50 = 224$

Hence from Equation 8 we have that

$$\hat{\pi}^+ = \frac{F^+}{n_1 n_2} = \frac{274}{340} = 0.806; \hat{\pi}^0 = \frac{F^0}{n_1 n_2} = \frac{16}{340} = 0.047 \text{ and } \hat{\pi}^- = \frac{F^-}{n_1 n_2} = \frac{50}{340} = 0.147.$$

Now $\pi^+ - \pi^-$ is the probability that a randomly selected subject from the population admitted into the hospital for hypertension stays longer less than the probability that he stays shorter than a randomly selected subject from the population admitted to a hospital for malaria. Its sample estimate is from Equation 9, $\hat{\pi}^+ - \hat{\pi}^- = 0.806 - 0.147 = 0.659$. The corresponding sample variance is from Equation 10

$$Var\left(\hat{\pi}^+ - \hat{\pi}^-\right) = \hat{\pi}^+ + \hat{\pi}^- - \frac{\left(\hat{\pi}^+ - \hat{\pi}^-\right)^2}{n_1 n_2} = 0.806 + 0.147 - \frac{(0.806 - 0.147)^2}{340} = 0.002$$

the test statistic of the null hypothesis, H_0 of Equations 11 or 12 that hypertension and malaria patients from the population admitted to a hospital for treatment have on the average equal median lengths of hospitalization for their illnesses is from Equation 13 with $\theta_0 = 0$

$$\chi^2 = \frac{n_1 n_2 \left(\hat{\pi}^+ - \hat{\pi}^-\right)^2}{\hat{\pi}^+ + \hat{\pi}^- - \left(\hat{\pi}^+ - \hat{\pi}^-\right)^2} = \frac{(0.659)^2}{0.002} = \frac{0.434}{0.002} = 217.00\,(P - value = 0.0000)$$

which with 1 degree of freedom is highly statistically significant at the 5 percent significance level $\left(\chi^2_{0.95;1} = 3.841\right)$. Hence we would reject the null hypothesis and conclude that hypertension and malaria patients in the population do not have equal median lengths of hospitalization. Having rejected the null hypothesis H_0 of Equation 11 and 12 that hypertension and malaria patients in the population have different median lengths of hospitalization, one may now proceed to further determine which of the population of patients, whether hypertension or malaria patients in the population have median length of hospitalization that is significantly different from the overall or common median length of hospitalization for both hypertension and malaria patients pooled together and hence may have led to the rejection of the initial null hypothesis of equal median lengths of hospitalization for hypertension and malaria patients in a population. To do this, we pooled together the two samples of hypertension and malaria patients into one combined sample of size $n_1 + n_2 = 20 + 17 = 37$ patients and determine the common sample median of the combined sample observations which is here found to be m=7 days. Now using m=7days in Equation 15 we calculate the values of u_{ix_1} *and* u_{ix_2} for the sample data of Table 1.The results are shown in Table 3.

Table 3: Values of u_{ix_1} *and* u_{ix_2} of Equation 15 for the sample data of Table 1 with m=7days

(i)	x_{i1} (hypertension)	u_{ix_1}	(i)	x_{i2} (malaria)	u_{ix_2}
1	7	0	1	7	0
2	4	-1	2	11	1
3	9	1	3	2	-1
4	18	1	4	7	0
5	5	-1	5	3	-1
6	17	1	6	7	0
7	17	1	7	3	-1
8	6	-1	8	1	-1
9	5	-1	9	7	0
10	13	1	10	8	-1
11	16	1	11	5	-1
12	10	1	12	3	0
13	12	1	13	3	1
14	7	0	14	2	-1
15	5	-1	15	5	-1
16	12	1	16	3	-1
17	13	1	17	4	-1
18	9	1			
19	4	-1			
20	10	1			

The summary values of u_{ix_1} *and* u_{ix_2} **and other statistics are presented in Table 4.**

Table 4: Summary values of u_{ix_1} *and* u_{ix_2} (Equation 15) of Table 3 and other statistics.

Index

Illness(i)	$F_{\cdot i}^+$	$F_{\cdot i}^0$	$F_{\cdot i}^-$	W $\left(F_{\cdot i}^+ - F_{\cdot i}^-\right)$	n	$\hat{\pi}_{\cdot i}^+$	$\hat{\pi}_{\cdot i}^0$	$\hat{\pi}_{\cdot i}^-$	$\hat{\pi}_{\cdot i}^+ - \hat{\pi}_{\cdot i}^-$ (24)	$Var\left(\hat{\pi}_{\cdot i}^+ - \hat{\pi}_{\cdot i}^-\right)$ (25)	$\chi_{\cdot i}^2(27)$
Hypertension (x_{i1})	12	2	6	6	20	0.6	0.1	0.3	0.30	.041	2.195
Malaria (x_{i2})	2	4	11	-9	17	.118	.235	.647	-0.529	.029	9.655

It is seen from Table(4)that, the Chi-square value for testing the null hypothesis H_0 of Equation (26) that the median length of hospitalization of hypertension patients in the population is equal to the median length of hospitalization of both hypertension and malaria patients in the population when pooled together as one population, is

$\chi^2 = 2.195(P-value = 4.345)$ which with 1 degree of freedom is not statistically significant at the 5 percent significance level $\left(\chi^2_{0.95;1} = 3.841\right)$ leading to the non-rejection of the null hypothesis. On the other hand, the Chi-Square value for testing the same null hypothesis with respect to malaria patients, that is that the median length of hospitalization of malaria patients in the population is the same as the median length of hospitalization of the combined or pooled population of hypertension and malaria patients when combined and treated as one population is $\chi^2 = 9.655(P-value = 0.2134)$ which with 1 degree of freedom is statistically significant at the 5 percent significance level $\left(\chi^2_{0.95;1} = 3.841\right)$ leading to the rejection of the null hypothesis. We may therefore conclude that the median length of hospitalization of malaria patients is statistically different from the median length of hospitalization of both hypertension and malaria patients in the sampled population and may hence be responsible for the rejection of the initial null hypothesis H0 of Equation 11 or 12 of equal population median lengths of hospitalization of the two types of patients in the sampled population. In fact, note that the median length of hospitalization of malaria patients is here estimated from the same data of Table (1) as $m_2 = 4days$ which is much smaller than the same estimate of median length of hospitalization of hypertension patients of Table (1) as $m_1 = 9.5days$ which is closer to $m = 7days$ obtained for the combined sample of malaria and hypertension patients of Table (1).The above more detailed analysis and conclusions cannot be readily reached using only the ordinary median test for two samples. It would be instructive to compare the proposed modified ties adjusted median test with the usual ordinary unmodified ties unadjusted median test for two sample. To do this we note that the common median of the pooled sample of hospitalization days of malaria and hypertension patients in the population is m=7days.Now altogether 6 observations have the same value of 7days as the common median m=7days,4 for malaria and 2 for hypertension. Hence these 6 observations are discarded in further analysis (Ebuh and Oyeka,2012) given as effective total sample size of m=37-6=31. Using the sample data of Table (1) with adjusted sample size of m=31 we obtain the fourfold table (Table 5) partitioning with m=7days, the patients of Table (1) by type of illness into those patients whose length of hospitalization is more than m=7 days and those patients whose length of hospitalization is less than m=7 days. The results are shown in Table (5).

Table (5):2 x 2 table for the sample data of Table (1) with m=7 days.
Types of illness

Length of hospitalization	Hypertension	Malaria	Total$(n_{i.})$
More than m=7days	$12(F_1^+ = n_{11})$	$2(F_2^+ = n_{12})$	$14(n_1 = F^+)$
Less than m=7days	$6(F_1^- = n_{21})$	$11(F_1^- = n_{22})$	$17(n_2 = F^-)$
Total$(n_{.j})$	$18(n_1)$	$13(n_2)$	$31(n = n)$

Note that the cell frequencies of Table 5 is exactly the same as the cell frequencies of Table 4 without adjustment for tied observations, that is if we had excluded from Table 4 all the patients in the sample whose length of hospitalization for their illness is equal to the common sample median days of hospitalization m=7days.Hence Table (5) could as well have been obtained from Table 4 by deleting all the F^0 values. Now the Chi-square test statistic for testing the null hypothesis H0 of independence or no association between cell frequencies in a fourfold or 2x2 frequency table is (Ebuh and Oyeka, 2012) using the notations of Table (5).

$$\chi^2 = \frac{n\left(n_{11}n_{22} - n_{12}n_{21}\right)^2}{n_1 n_2 n_{.1} n_{.2}} \quad (31)$$

Which under the null hypothesis H0 has approximately 1 degree of freedom for sufficiently large sample size n. Using the sample observation (length of hospitalization of patients in days) in Table (5) with Equation 31 we have

$$\chi^2 = \frac{31(12\times 11 - 2\times 6)^2}{14\times 17\times 18\times 13} = 8.016(P-value = 0.1683)$$

which with 1 degree of freedom is also statistically significant at the 5 percent significance level $\left(\chi^2_{0.95;1} = 3.841\right)$, again leading to the rejection of the null hypothesis that hypertension and malaria patients in the sampled population have equal lengths of hospitalization for their illnesses. However since the Chi-square value of $\chi^2 = 217.00$ obtained using the proposed modified ties adjusted median test for two samples is much larger than the Chi-square value of $\chi^2 = 8.016$ obtained using the usual ordinary unmodified ties unadjusted two sample median test, the proposed method is likely to correctly reject a false null hypothesis more often and hence is more powerful than the ordinary median test when used to analyze the same sample observations. It would also be instructive to compare the results obtained using the proposed method with the results that would be obtained if we had used the Mann-Whitney U-test (Gibbons,1992) to analyze the same sample data. To use the Mann-Whitney U-test we first pool the two sample observations from populations x_1 *and* x_2 into one combined sample of total sample size $n_1 + n_2$.The combined or pooled sample observation are now ranked from smallest value, say, assigned the rank 1 to the largest value assigned the rank

$n_1 + n_2$.All tied observations in the combined or pooled sample are assigned their mean ranks. Let $R_{.1}$ be the sum of the ranks assigned to sample observation from population x_1 and $R_{.2}$ be the sum of the ranks assigned to sample observation from population x_2.In the combined ranking of these sample observations. The Mann-Whitney U-test statistic is (Gibbons,1992)

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_{.1} \quad (32)$$

OR equivalently

$$U' = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_{.2} \quad (33)$$

With mean and variance as respectively

$$E(u) = \mu = \frac{n_1 n_2}{2}; Var(u) = \frac{n_1 n_2(n_1+n_2+1)}{12} \quad (34)$$

The Mann-Whitney U-test statistic is

$$\chi^2 = \frac{(u-\mu)}{Var(u)} = \frac{u - \frac{n_1 n_2}{2}}{\frac{n_1 n_2(n_1+n_2+1)}{12}} \quad (35)$$

Which under the null hypothesis H0 has approximately the Chi-square distribution with 1 degree of freedom for sufficiently large sample size $n = n_1 + n_2 = 20 + 17 = 37$ values of the lengths of hospitalization of the combined or pooled samples of hypertension and malaria patients from the smallest value of 1 day ranked 1 to the largest value of 18 days ranked 37 assigning all tied lengths of hospitalization their mean ranks. The sums of the ranks assigned to sample observations from populations x_1 (hypertension) and x_2 (malaria)are $R_{.1} = 494 \; and \; R_{.2} = 209$ respectively. Using this results with $n_1 = 20, n_2 = 17, \; and \; R_{.1} = 494.$ In equations 32 and 34,we now obtained the values of the U statistic, its mean and variance as respectively

$$u = (20)(17) + \frac{20(21)}{2} - 494 - 56; E(u) = \mu = \frac{(20)(17)}{2} = 170;$$

$$Var(u) = (20)(17)\frac{(20+17+1)}{12} = \frac{340(38)}{12} = 1076.67$$

Using these value in Equation 35 to obtain the Mann-Whitney U Test Statistic

$$\chi^2 = \frac{(56-170)^2}{1076.67} = \frac{12996}{1076.67} = 12.071(P - value = 1.027)$$

Which with 1 degree of freedom is also statistically significant at the 5 percent significance level again leading to the rejection of the null hypothesis H0 of equal median lengths of hospitalization of hypertension and malaria patients in the sampled population. However, the relative sizes of the calculated Chi-square values of $\chi^2 = 12.071$ for the Mann-Whitney U test and Chi-square value of $\chi^2 = 217.00$ for the proposed modified ties adjusted test statistic which is much larger suggest that the Mann-Whitney U test is likely going to lead to an acceptance of a false null hypothesis(Type II Error)more often than the modified intrinsically ties adjusted two samples median test and hence is less powerful than the proposed method when used to analyze the same sample data.

Conclusions

We have in this paper proposed, developed and presented a non-parametric statistical method for the analysis of two sample data that intrinsically and structurally adjusts the test statistic for the possible presence of tied observation between the sampled populations thereby obviating the need to require these populations to be continuous or even numeric as is usually the case with some other methods used in these types of analysis. The populations may be measurements on as low as the ordinal scale and need not be continuous or numeric. Test statistics are provided including the test statistics to use in determining which of the sample populations may have possibly led to the rejection of an original null hypothesis, if rejected of equal population medians, a procedure that is not possible with some existing methods for two sample data analysis. The proposed methods are illustrated with some sample data and shown to be at least as powerful and efficient as some existing non-parametric statistical methods that could equivalently be used for the same purpose including the Mann-Whitney U test and the ordinary ties unadjusted median test for two samples.

Conflict of interest:

No conflict of interest.

Acknowledgement

The authors wish to thank the referees for their worthwhile comments and suggestions.

References

1. Corder, G. W. and Foreman, D. I. (2014). Nonparametric Statistics: A step-by-step Approach, Wiley. ISBN 978-1118840313.
2. Gibbons, J. D. (1993). Nonparametric Statistical; An Introduction, Newbury Park, Sage Publication. Pp. 180-220. Gibbons, J. D. and Chakraborti, S. (2003). Nonparametric Statistical Inference. Fourth Edition, Revised and Expanded, Marcel Dekker, New York.
3. Ebuh, G. U. & Oyeka, I. C. A. (2012). Statistical Comparison of Eight Alternative Methods for the Analysis of Paired Sample Data with Applications. Open Journal of Statistics (OJS), 2(3), 328-345.
4. Friedlin, B. and Gastwirth, J. L. (2000). Should the median test be retired from general use? The American Statistician, 54, 161-164.
5. 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 Analysing Paired Data Intrinsically Adjusted for Ties. Global Journal of Mathematics and Statistics, India. 1(1), 1-6
6. Oyeka Cyprain Anaene, Chigozie Edson Utazi, Chinwe Rosemary Nwosu, Godday Uwawunkonye Ebuh, Precious Ada Ikpegbu, Happiness Obiorah-Ilouno and Christian Chukwuemeka Nwankwo. 2010. “A Statistical Comparison of Test Scores: A Non-Parametric Approach”. Journal of Mathematical Sciences. Vol 21, No.1 (2010) Pgs

77-87.

7. Oyeka I. C. A., Obiora-Ilouno H., Ebuh G.U., Nwankwo C.C. and Umeh E.U. (2011) "Pairwise Comparison of Three Populations: A Nonparametric Method. Journal of Basic Physical Research. ISSN;2141-8403-2141-8411 online. Vol 2, No.1, PP 25-35, Nigeria.
8. Oyeka, C. A. (2013). An Introduction to Applied Statistical Methods.Ninth Edition, Nobern Avocation Publishing Company, Enugu.
9. Siegel S. (1988): Non-parametric Statistics for the Behavioural Sciences. McGraw-Hill Kogakusha Ltd.,
10. Tokyo, 399pp
11. Siegel, S. and Castellan, N. J. Jr. (1988). Nonparametric Statistics for the Behavioral Sciences.New York, McGraw-Hill.