

Volume 10, Issue 11, November 2024,
Publish Date: 22-11-2024
Doi https://doi.org/10.55640/ijmsdh-10-11-08

International Journal of Medical Science and Dental Health (Open Access)

PREDICTING FERTILITY STATUS BY ADJUSTING FOR TIES IN A MATCHED SAMPLED POPULATION IN EBONYI STATE USING MEDIAN TEST

OKEH, UCHECHUKWU MARIUS[1,2], 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.

ABSTRACT

Background and Aim of the study: The use of nonparametric methods in statistical procedures occurs when samples drawn fails to satisfy the necessary assumptions of continuity and normality required for their use by their parametric counterpart. We proposed and developed ties adjusted median test for determining desired fertility goals in a paired sampled population. This is an improved alternative to the ordinary Sign test and the Wilcoxon Signed Rank test. **Methodology:** The proposed method adapted the method of extended median test for matched samples but in this case adjusted for possible tied observations in the data where observations or scores may be measurements on as low as the ordinal scale. Chi-square test statistic of null hypothesis of equality of population or treatment medians was used. **Results:** The data for illustrating the proposed method was drawn from the actual and desired family sizes of a random sample of women from a certain community. Applying the data in testing the null hypothesis that women from the community sampled do not differ in the actual and desired number of children showed that Using the same data, sign test gave P-value=0.0193 while Wilcoxon Signed Rank test gave P-value=0.0139 **Conclusions:** Results showed that the proposed method is more powerful than the ordinary Sign test and the Wilcoxon Signed Rank test at 5% significant level and hence is likely to have a higher probability of correctly rejecting a false null hypothesis.

KEYWORDS: extended median test, sign test, paired samples, Wilcoxon Signed Rank Sum test, chi-square test.

INTRODUCTION

The paired sample parametric test may be used to analyze paired sample data and test a desired null hypothesis if the two populations from which the samples are drawn satisfy the necessary assumptions of continuity and normality [1,2]. If these assumptions are not satisfied, then a preferred nonparametric test to use would be either the Sign test or the Wilcoxon Signed Rank test provided there are no tied or only few tied observations in the data [3,4]. If there are only few tied observations, these observations may be dropped and the sample size reduced accordingly in subsequent analysis or the tied observations may be assigned their mean ranks in the case of the Wilcoxon Signed Rank test [5,6]. If, however the number of ties is not few, then these nonparametric tests may not be properly used without adjustments for ties [7,8]. This is because too many ties in the data, if not adjusted for, may seriously compromise the power of the test and may lead to unreliable conclusions. We here propose a tie adjusted median test for this purpose based on the extended median test for matched samples [1,9,10,11].

METHODOLOGY

Suppose we have a random sample of 'r' subjects or candidates observed at two points in time or space or under two experimental conditions or treatments or assessed by two judges or instructors. Let x_{ij} be the score by the ith subject under treatment j for i=1, 2,..,r and j=1,2.These observations or scores may be measurements on as low as the ordinal scale. Let M_i be the median of the two observations on the ith subject. Note that actually M_i is the average of the two observations on the ith subject, that is, $M_i = (x_{i1} + x_{i2})/2$, for i=1, 2,….,r.

Now let

$$u_{ij} = \begin{cases} 1, if \ \ x_{ij} > M_i \\ 0, if \ \ x_{ij} = M_i \\ -1, if \ \ x_{ij} < M_i \end{cases} \qquad 1$$

For i=1, 2., r; j=1,2.

Note that, Equation 1 is equivalent to saying that u_{ij} assumes the value 1,0, -1 if one of the observations on the ith subject is greater than, equal to, or less than the other observation. Thus, the present method is in this respect similar to the approach used in the ordinary Sign test except that provision has now been made for the possible presence of ties in the data. Let t_j^+, t_j^0 *and* t_j^- be respectively the number of 1s,0s and -1s in the experimental condition or treatment j for all i=1,2,….,r ;j=1,2.Note that $t_j^0 = r - t_j^+ - t_j^-$. Also let t^+, t^0 *and* t^- be respectively the total number of 1s,0s and -1s in u_{ij} .That is

$$t^+ = \sum_{j=1}^{2} t_j^+ ; t^- = \sum_{j=1}^{2} t_j^- ; t^0 = \sum_{j=1}^{2} t_j^0 = 2r - t^+ - t^- \qquad 2$$

The proportions of 1s, -1s and 0s at the jth level of treatment or experimental condition, that is the jth level(j=1,2) of the responses for all the subjects are respectively

$$P_j^+ = \frac{t_j^+}{r} ; P_j^- = \frac{t_j^-}{r} ; P_j^0 = \frac{t_j^0}{r} = 1 - P_j^+ - P_j^- . \qquad 3$$

The overall proportions of 1s, -1s and 0s are respectively

$$P^{+}=\frac{t^{+}}{2r};P^{-}=\frac{t^{-}}{2r};P^{0}=\frac{t^{0}}{2r}=1-P^{+}-P^{-}. \qquad 4$$

The observed number of 1s, -1s and 0s at the jth level of response are respectively

$$O_{1j}=t_j^{+};O_{2j}=t_j^{-};O_{3j}=t_j^{0}=r-t_j^{+}-t_j^{-} \qquad 5$$

Under the null hypothesis of no difference between the proportions of responses for the two treatment levels, that is the null hypothesis of equal population medians, the corresponding expected frequencies are

$$E_{1j}=\frac{rt^{+}}{2r};E_{2j}=\frac{rt^{-}}{2r};E_{3j}=\frac{rt^{0}}{2r}=\frac{r\left(2r-t^{+}-t^{-}\right)}{2r} \qquad 6$$

To test the null hypothesis of equal population medians, we use the test statistic

$\chi^2=\sum_{j=1}^{2}\sum_{i=1}^{3}\frac{\left(O_{ij}-E_{ij}\right)^2}{E_{ij}}$ which under the null hypothesis H0 has approximately the Chi-square distribution with 2 degrees of freedom, substituting Equations 5 and 6 in the above expression yields

$$\chi^2=\frac{\sum_{j=1}^{2}\left(t_j^{+}-\frac{rt^{+}}{2r}\right)^2}{\frac{rt^{+}}{2r}}+\frac{\sum_{j=1}^{2}\left(t_j^{-}-\frac{rt^{-}}{2r}\right)^2}{\frac{rt^{-}}{2r}}+\frac{\sum_{j=1}^{2}\left((r-t_j^{+}-t_j^{-})-\frac{r(2r-t^{+}-t^{-})}{2r}\right)^2}{\frac{r(2r-t^{+}-t^{-})}{2r}}$$

That is

$$\chi^2=2\left(\frac{\sum_{j=1}^{2}\left(t_j^{+}-\frac{t^{+}}{2}\right)^2}{t^{+}}+\frac{\sum_{j=1}^{2}\left(t_j^{-}-\frac{t^{-}}{2}\right)^2}{t^{-}}+\frac{\sum_{j=1}^{2}\left(\left(t_j^{+}-\frac{t^{+}}{2}\right)+\left(t_j^{-}-\frac{t^{-}}{2}\right)\right)^2}{2r-t^{+}-t^{-}}\right)$$

Which when further simplified yields

$$\chi^2=\frac{2\left(t^{+}\left(2r-t^{-}\right)\sum_{j=1}^{2}\left(t_j^{+}-\frac{t^{+}}{2}\right)^2+t^{+}\left(2r-t^{+}\right)\sum_{j=1}^{2}\left(t_j^{-}-\frac{t^{-}}{2}\right)^2+2t^{+}t^{-}\sum_{j=1}^{2}\left(t_j^{+}-\frac{t^{+}}{2}\right)\left(t_j^{-}-\frac{t^{-}}{2}\right)\right)}{t^{+}t^{-}\left(2r-t^{+}-t^{-}\right)} \qquad 7$$

Or when expressed in terms of the proportions in equations 3 and 4 becomes

$$\chi^2 = \frac{r\left(P^-\left(1-P^-\right)\sum_{j=1}^{2}\left(P_j^+-P^+\right)^2 + P^+\left(1-P^+\right)\sum_{j=1}^{2}\left(P_j^--P^-\right)^2 + 2P^+P^-\sum_{j=1}^{2}\left(P_j^+-P^+\right)\left(P_j^--P^-\right)\right)}{P^+P^-\left(1-P^+-P^-\right)} \quad 8$$

Which has the Chi-square distribution with (3-1) (2-1) =2degrees of freedom.

Equation 8 may be simplified in an easier way to obtain the computational form as

$$\chi^2 = \frac{r\left(P^-\left(1-P^-\right)\left(P_1^+-P_2^+\right)^2 + P^+\left(1-P^+\right)\left(P_1^--P_2^-\right)^2 + 2P^+P^-\left(P_1^+-P_2^+\right)\left(P_1^--P_2^-\right)\right)}{2P^+P^-\left(1-P^+-P^-\right)} \quad 9$$

Note that actually $P_1^+ = P_2^-, P_1^- = P_2^+ \ and \ P^+ = P^-$. Hence Equation 9 may be further simplified to

$$\chi^2 = \frac{r\left(P_1^+-P_2^+\right)^2}{P^+} = \frac{r\left(P_1^--P_2^-\right)^2}{P^-} \quad 10$$

Or equivalently

$$\chi^2 = \frac{2r\left(P_1^+-P_2^+\right)^2}{2P^+} = \frac{2r\left(P_1^--P_2^-\right)^2}{2P^-} \quad 11$$

The null hypothesis H0 of equal population or treatment medians is rejected at the α level of significance if

$$\chi^2 \geq \chi^2_{1-\alpha;2} \quad 12$$

Otherwise H0 is accepted.

ILLUSTRATIVE EXAMPLE

A family planning consultant is interested in determining the extent to which Women in a certain community in Ebonyi State are able to achieve their desired fertility goals. She selected a random sample of women from the community and asked them to state their desired and actual or achieved family sizes and obtained the results shown in table 1.

Table 1 *(Actual and Desired Family sizes of a Random Sample of women from a certain community.)*

Women	**Actual no of children**	**Desired no of children**	$M_i = (x_{i1}+x_{i2})/2$	**Score for Treatment**	**Total**	**Proportion**

I	x_{i1}	x_{i2}		u_{i1}	u_{i2}	(t)	(p)
1	4	5	4.5	-1	1		
2	1	5	3	-1	1		
3	6	5	5.5	1	-1		
4	1	6	3.5	-1	1		
5	7	5	6	1	1		
6	1	9	5	-1	1		
7	4	4	4	0	0		
8	2	6	4	-1	1		
9	8	8	8	0	0		
10	5	5	5	0	0		
11	4	4	4	0	0		
12	4	5	4.5	-1	1		
13	5	6	5.5	-1	1		
14	5	6	5.5	-1	1		
15	4	4	4	0	0		
16	4	6	5	-1	1		
17	5	6	5.5	-1	1		
t_j^+				2	10	12	
t_j^-				10	2	12	
t_j^0				5	5	10	
Total				17	17	34	
p_j^+				$\frac{2}{17}=0.118$	$\frac{10}{17}=0.588$		$\frac{12}{34}\,p=0.353p^+$
p_j^-				$\frac{10}{17}=0.588$	$\frac{2}{17}=0.118$		$\frac{12}{34}\,p=0.353p^-$
p_j^0				$\frac{5}{17}=0.294$	$\frac{5}{17}=0.294$		$\frac{10}{34}\,p=0.294p^0$

The median ''family size'' M_i is calculated as the average of her actual and desired number of children and is shown in the fourth column of Table 1.These values are used in Equation 1 to obtain the u_{ij} values shown in columns 5 and 6 of Table 1.These scores are now used in Equations 3 and 4 to obtain the proportions shown at the bottom of Table 1.These proportions are now used in Equation 10 to test the null hypothesis that women from the community sampled do not differ in the actual and desired number of children they have giving

$\chi^2 = \frac{(17)(0.118-0.588)^2}{0.353} = \frac{3.755}{0.353} = 10.637(P-value = 0.0048)$ which with 2 degrees of freedom is statistically significant at 5% percent level of significant ($\chi^2_{0.99;2} = 9.210$).Hence we may conclude that actual and desired family sizes of women in the sampled community do in fact differ. We may now use the ordinary sign test to reanalyze the data for comparative purposes. Note that five women achieved their desired family sizes, meaning that there are five ties in the data hence the effective sample size to use for the sign test is n=17-5=12.Also since there is only x=2+signs, we calculate assuming P=0.5 under the usual null hypothesis for the sign test,

$P(x \le 2) = \sum_{x=0}^{2} \binom{12}{2} (0.5)^{12} = (1+12+.66)(0.000244) = 0.0193$ which is statistically significant at the 5 percent but not at the 1 percent level. Also, the attained significance probability (P-value=0.0193) under the sign test is higher than the corresponding value (0.0048) under the proposed method indicating that the sign test is likely to lead to an acceptance of a false null hypothesis (Type II Error) more often and hence less powerful than the proposed method. It would also be instructive to compare the proposed method with the Wilcoxon Signed Rank test. To do this we take the differences 'di' between the actual and desired family sizes, take their absolute values and rank only the 12 absolute values with non-zero differences. The ranks assigned to the absolute values with positive differences which are fewer in number are 3.5 and 7.5 summing to 11.0. The Wilcoxon Signed Rank test statistic is then

$$Z = \frac{T-\mu}{\sqrt{Var(T)}}, where\ T = 3.5+7.5 = 11.0; \mu = \frac{12(13)}{4} = 39.0, Var(T) = \frac{12(13)(25)}{24} = 162.5.$$

Hence

$$Z = \frac{11.0-39.0}{\sqrt{162.5}} = \frac{-28.0}{12.748} = -2.20(P-value = 0.0139)\ which\ is\ also\ statistically\ significant\ at\ the\ 5\ percent$$

But not at the 1 percent significant level indicating that the proposed method may also be more powerful than the Wilcoxon Signed Rank test.

CONCLUSION

A non-parametric statistical method has here been developed for the analysis of paired sample data given the fact that the two populations from which the samples are drawn failed to satisfy the necessary assumptions of continuity and normality. We therefore here adapted the method used in the extended median test for matched samples but in this case adjusted for possible tied observations in the data. It is an improved alternative to the ordinary Sign test and the Wilcoxon Signed Rank test. The proposed method which resulted to the Chi-square test of null hypothesis of equality of population or treatment medians is illustrated with some data on the actual and desired family sizes of a random sample of women from a certain community and the result was compared with results obtained when Sign test and the Wilcoxon Signed Rank test was applied on the same data. Meanwhile, tied observations seen in the data was discarded as a prerequisite for using Sign test or the Wilcoxon Signed Rank test. Results obtained showed that the proposed method is more powerful than the ordinary Sign test and the Wilcoxon Signed Rank test at 5% significant level since the attained significant probability (P-value=0.0193) under the sign test and the significant probability (P-value=0.0139) under the Wilcoxon

Signed Rank test are all higher than the corresponding significant probability (P-value=0.0048) under the proposed method indicating that the sign test and the Wilcoxon Signed Rank test are likely to lead to an acceptance of a false null hypothesis (Type II Error) more often and hence are less powerful than the proposed method.

DISCUSSION

We have in this paper investigated a situation where the use of parametric statistical method failed due to the inability of the data of interest to attain the necessary assumptions of continuity and normality. This situation called for the application a preferred nonparametric statistical test where in this case only few tied observations were seen in the data presented and those tied observations were dropped and the sample size reduced accordingly in subsequent analysis. In other similar situations, the tied observations may be assigned their mean ranks. In this paper, developed a tie adjusted median test for paired samples, based on the extended median test for matched samples. We illustrated the proposed method using a sampled data and compared the result obtained with results obtained using the Sign test and Wilcoxon Signed Rank test on the same data. Results showed the proposed ties adjusted median test is more powerful than the ordinary Sign and Wilcoxon Signed Rank tests. Similar work indicating that the proposed modified extended median test is more powerful than the ordinary extended median test was carried out [11]. Recall that this paper was also based on ordinary extended median test. Further study is recommended where there is need to adjust for ties in Sign and Wilcoxon Signed Rank tests and then compare their performance with ties adjusted median test using any data of interest. This is to determine their statistical power.

REFERENCES

1. Gibbons, J. D. (1993). Nonparametric Statistical; An Introduction, Newbury Park, Sage Publication. Pp. 180-220.
2. Oyeka, C. A. (2013). An Introduction to Applied Statistical Methods. Ninth Edition, Nobern Avocation Publishing Company, Enugu.
3. 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.
4. 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.
5. Corder, G. W. and Foreman, D. I. (2014). Nonparametric Statistics: A step-by-step Approach, Wiley. ISBN 978-1118840313.
6. Siegel, S. and Castellan, N. J. Jr. (1988). Nonparametric Statistics for the Behavioral Sciences. New York, McGraw-Hill.
7. Gibbons, J. D. and Chakraborti, S. (2003). Nonparametric Statistical Inference. Fourth Edition, Revised and Expanded, Marcel Dekker, New York.
8. Friedlin, B. and Gastwirth, J. L. (2000). Should the median test be retired from general use? The American Statistician, 54, 161-164.
9. Gart, J. J. (1963). A Median Test with Sequential Application. Biometrika, vol. 50, pp. 55-62.
10. Afuecheta, E. O., Oyeka, I. C. A, Ebuh, G. U., & Nnanatu, C. C. (2012). Modified Median Test Intrinsically Adjusted for Ties. Journal of Basic Physical Research, 3, 30-34.
11. Oyeka I. C. A. and Uzuke C. A. (2011). Modified extended median test. African Journal of Mathematics and Computer Science Research Vol. 4(1), pp. 27-31.