Non-Parametric Test for Ordered Medians: The Jonckheere Terpstra Test

Arif Ali[1], Abdur Rasheed[1,*], Afaq Ahmed Siddiqui[2], Maliha Naseer[3], Saba Wasim[3] and Waseem Akhtar[4]

[1]*Department of Research, Dow University of Health Sciences, Karachi, Pakistan*

[2]*Department of Pharmacy, University of Karachi, Karachi, Pakistan*

[3]*Department of Community Health Sciences, Aga University, Karachi, Pakistan*

[4]*Department of Radiology, Aga Khan University, Karachi, Pakistan*

Abstract: In clinical trials, sample size is usually lesser as compared to other epidemiological studies to make it more feasible and cost effective. Small sizes of such trials discourage the use of parametric test due to violation of the assumption under which they are applicable. Therefore, the use of nonparametric test is substantial in clinical trials to test two or more independent samples. The Kruskal-Wallis h test is an alternative to one-way ANOVA and can be used to identify significant differences among different populations. When we have several independent samples and assumed to be arranged orderly, Jonckheere Terpstra test is a best choice to compare population medians instead of means. For the application of Jonckheere Terpstra test the data from the study of cleaning methods for ultrasound probes are used. The Jonckheere Terpstra test is recommended over Kruskal-Wallis h test as it compares and provides significant difference between more than two population medians when they arranged in order. Therefore, the aim of this research paper was to explore the use and significance of Jonckheere-Terpstra test with the use of practical example.

Keywords: Jonckheere Terpstra test, non parametric test, comparison of medians.

INTRODUCTION

Wolfowitz [1] was first who introduced the term nonparametric that encompasses techniques that do not depend on data that belong to a specific distribution [2].

Nonparametric tests are also called distribution-free tests [3]. These nonparametric tests do not require normality and homogeneity of variance assumptions. Unlike to parametric tests, nonparametric tests compare medians, therefore, in the presence of one or two outliers their effect is negated.

In clinical trials, sample size is usually less smaller to make it more feasible and cost effective [4]. Due to small sample size the assumptions of normality and about the necessity of having homogeneity of variance within groups are not met for parametric tests (such as t-test and one-way ANOVA). Therefore, in such situations it is appropriate to use non parametric tests such as for two independent samples the Wilcoxon rank sum test or Wilcoxon-Mann-Whitney, and for more than two independent samples the Kruskal-Wallis h test and Jonckheere Terpstra test. The aim of this research paper was to explore the use and significance of Jonckheere-Terpstra test with the help of practical example when the Kruskal-Wallis h test was not appropriate choice.

The Kruskal–Wallis h test is an extension of the Mann–Whitney test and used for more than two independent samples. The Kruskal–Wallis h is the nonparametric test an alternative to one-way ANOVA and can be used to identify significant differences among different population groups with the null hypothesis of having no difference among different groups and the alternative hypothesis of difference between at least two of them. Instead of comparing population means, this test compares population medians.

Most medical studies continue to apply Kruskal–Wallis h test with relevant lack of statistical power. For example a clinical prospective study [5] was conducted for the evaluation of endoscopic resected specimen and tumor size. In that study the investigation was performed to identify whether the true tumor diameter gastrointestinal cancer specimen measure by flexible endoscopy is relevant of testing whether specimen is correctly attached after endoscopic submucosal dissection resection furthermore, whether, the size variations depend on the endoscopist who attached the specimen. Descriptive statistics for resected specimen diameters of three endoscopists and also recorded tumor sizes were reported. Furthermore, Kruskal Wallis h- test was performed to identify the significant

*Address correspondence to this author at the Department of Research, Dow University of Health Sciences, Karachi, Pakistan;
E-mail: abdur.rasheed@duhs.edu.pk

difference of resected specimen diameters of three endoscopists and same as for tumor sizes. Kruskal Wallis h- test provided significant p-values 0.040 and 0.031 for both specimen diameters and tumor sizes. But authors did not try to identify order of medians in which specimen diameters and tumor sizes were differ among three groups or which group have higher median diameters or tumor size than other group.

Another study [6] was found to determine the relationship between best practice tariff (BPT) achievement and important patient outcomes and whether the BPT could predict these independently of other validated predictors. Retrospective review was carried on 516 patient episodes. Four outcomes were defined: i) 30-day mortality, ii) 365-day mortality, iii) postoperative length of stay on trauma ward (LOS-T), and iv) total post-operative hospital LOS (LOS-H). Patient episodes were divided in to three groups as follows: 1) group 1, pre-BPT, 2) group 2, BPT achievers, 3) group 3, BPT fails. In addition to other findings, It was reported that (LOS-T) was significantly differ among three groups (Kruskal Wallis h- test; P-value =0.005) but no further analysis were carried on to obtained statistical evidence about the order of medians of length of stay on trauma wards among three different groups based on BPT.

Therefore, there are situations in which order of treatments may be substantial and a test with the alternative hypothesis based on ordered population medians is required. For instance, the alternative hypothesis could be as follows: population median1 ≤ population median2 ≤ population median3 ≤ population median4 ≤…..≤ population median n. This is a one-tail test, and reversing the inequalities gives an analogous test in the opposite tail. Now to deal such situations where the alternative hypothesis involves distributional characteristics with inequalities Jonckheere-Terpstra is quite appropriate. Jonckheere-Terpstra test tests the null hypothesis that the distribution of the response variable does not differ among classes. It is developed to detect the alternative hypothesis of ordered class differences, which can be shown as $\eta_1 \leq \eta_2 \leq \ldots\ldots \leq \eta_T$ (or $\eta_1 \geq \eta_2 \geq \ldots\ldots \geq \eta_T$) with at least one of the inequalities being strict. Jonckheere-Terpstra test requires independent samples (grouping variable) be orderly arranged [7].

Test statistic T_{JT} is defined as:

$$J = \sum U_{xy}$$

Where U_{xy} is the number of observations in group y that are greater than each observation in group x

The standardized test statistic is computed as

$$Z = \frac{J - E(J)}{\sqrt{Var(J)}}$$

Here

$$E(J) = \frac{N^2 - \sum_{j=1}^{k} n_j^2}{4} \quad Var(J) = N^2(2N+3) - \sum_{j=1}^{k} n_j^2(2n_j+3)$$

Where,

N is the total number of observations in all groups, n_j is the observation in group j and j is the number of group and k is the total number of groups. For large N and individual n_j not to small, the distribution of test statistics Z is approximately standard normal.

METHODOLOGY

In this paper we considered the study of cleaning methods of ultrasound probes that was performed at the Radiology Department, the Aga Khan University Hospital, Karachi and Microbiology Department, JPMC, Karachi, from December 2006 to April 2007 [8].

Main sources of nosocomial infections are hospital or microorganism acquired from the hospital which lead to patient morbidity and mortality substantially [9, 10]. South-East Asian hospital reported the prevalence of nosocomial infection is 10% [10]. Extensive numbers of sonographic inspection are accomplished in tertiary care hospital as Ultrasonography suite is most frequently employed imaging modality.

It has been found in many studies that without proper and effective cleaning methods, Ultrasound (US) probes become the main path for spreading the pathologic organism from patient to patient [11-14].

In that study a total of 75 culture swabs from ultrasound probes of different body parts of patients for sonographic inspection were used. Three different ultrasound probe cleaning techniques were employed, first techniques include sterilized paper towel while second and third techniques include 0.9% saline and swipe over with standard bath soap. First technique was applied on group A (n=25), second was applied on group B (n=25) and third technique was applied on group C (n=25). To observe the effectiveness of different cleaning methods in reducing bacterial counts

Table 1: Number of Bacterial Count before and after Three Cleaning Ultrasound Probes Methods

Tissue paper			Saline		Soap	
S. no.	Before	After	Before	After	Before	After
1	350	136	292	51	213	11
2	142	62	302	42	296	13
3	190	106	261	49	312	9
4	300	190	302	97	268	7
5	409	211	192	39	202	5
6	390	192	201	32	312	4
7	159	61	192	62	257	8
8	198	101	289	67	361	2
9	302	192	290	81	301	6
10	296	136	233	89	331	6
11	322	166	209	41	296	3
12	172	72	289	53	326	2
13	104	78	301	89	396	6
14	151	91	189	39	307	2
15	133	71	161	39	256	1
16	202	131	231	61	303	3
17	102	89	142	29	309	2
18	109	79	190	58	268	8
19	167	99	203	81	292	7
20	79	59	297	52	302	2
21	107	78	219	51	368	6
22	89	55	161	21	317	1
23	202	121	232	41	314	1
24	197	101	171	36	316	2
25	106	79	193	71	309	5

the calculation of the Colony Forming Unit (CFU) of bacterial counts were carried on the standard agar plate. Data are given in the Table **1**.

RESULTS

Data were analyzed using SPSS v 21.0. At the first instance tests of Kolmogorov Smirnov and Shapiro Wilk were performed for the checking of normality assumption, the observed p-values were significant and enough to violate the normality assumption. Consequently, in order to find out significant differences in bacterial counts, statistical analysis was carried on through non-parametric tests.

A total of 75 ultrasound probes underwent three different cleaning methods after performing ultrasound procedures on the patients. Before employing cleaning methods, median bacterial counts for groups A, B and C are 172, 219 and 307 respectively. After using cleaning methods, median bacterial counts were reduced to 99, 51 and 5, therefore maximum bacterial reduction (98%) was observed in group C where soap wipe was used and reduction in group A and B were 42% and 76% where sterilized paper towel and Sline (0.9%) were used respectively.

We first performed Wilcoxon Signed Rank test for before and after comparison of bacterial count in each cleaning method at 5% level of significance. From Wilcoxon Signed Rank test it was observed that the differences between bacterial counts before and after in all three different cleaning methods was significant with P-values < 0.001.

We also performed Kruskal-Wallis h test on the data of bacteria counts before and after employing three cleaning methods and significant differences were observed before and after employing three cleaning methods (P-values = 6.26 $X10^{-6}$ and $1.2443X10^{-13}$ respectively). Furthermore, the post-hoc test for Kruskal-Wallis h test was carried on, and the significant differences were noted in each pair of cleaning methods.

After reaching the conclusion that all three cleaning methods are significantly different, one may take interest in the order of median bacterial counts in which all three cleaning methods differ. Kruskal-Wallis H test does not provide any such information hence not a suitable choice to attain the insight of orders of median bacterial counts in which all three methods are differ. Furthermore, we did not have any statistical evidence to claim about the orders of median bacterial counts. To perform this task a non-parametric Jonckheere Terpstra test is the best choice. On the basis of above findings we developed following alternative hypothesis.

Group	Cleaning method
A	Tissue paper
B	0.9% Saline
C	Bath Soap

H_a: median $_A$ ≥ median $_B$ ≥ median $_C$.

We applied Jonckheere Terpstra test using SPSS v.21.0 and obtained the value of standardized test statistics -8.456 with P-value = $2.7686X10^{-17}$. This P-value was enough to support above mentioned alternative hypothesis.

As we discussed earlier that after applying cleaning method median bacterial count for group A is greater than both groups B and C. Jonckheere Terpstra test findings statistically confirmed that median bacterial count for group C is lesser than group B, which is lesser than A.

DISCUSSION

In two samples independent test (Mann-Whitney) when null hypothesis shows that both sample medians are equal to each other, the alternative hypothesis says that sample medians are not equal to each other (two tailed test). In other situations when the researcher sets the null hypothesis that one sample median is less than (or greater than) other sample median the alternative hypothesis would state that one sample median is greater than equal to (or less than equal to) other sample median (one tailed tests). Therefore, it obvious that when the comparison is to be made between two samples, their characteristics say medians can be tested in some order e.g., one sample median is less or greater than other sample median. But when one takes interest in comparison of more than two sample's distributional characteristics such as medians, usually comparison is made that all groups are same versus at least one pair of groups is not same hence no order of medians is involved is this comparison. A non-parametric test, Jonckheere-Terpstra is substantial to deal such instances where testing of ordered medians is of interest.

From the above example, we showed that, that Kruskal-Wallis test does not provide any evidence of order of median bacterial counts in which they differed, however, the Jonckheere-Terpstra does. The results pointed out that when the different cleaning methods and ordered median bacterial counts are relevant; the Jonckheere-Terpstra test is better suitable choice.

CONCLUSION

From this course of work, it is concluded that Jonckheere Terpstra test is recommended over Kruskal-Wallis h test as it compares and provides significant difference among population medians when they arranged in some order.

REFERENCES

[1] Wolfowitz J. Additive partition functions and a class of statistical hypotheses. The Annals of Mathematical Statistics 1942; 13(3): 247-79. http://dx.doi.org/10.1214/aoms/1177731566

[2] Sprent P, Smeeton NC. Applied nonparametric statistical methods: CRC Press 2007.

[3] Altman DG, Bland JM. Parametric v non-parametric methods for data analysis. BMJ 2009; 338.

[4] Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. JAMA 1994; 272(2): 122-4. http://dx.doi.org/10.1001/jama.1994.03520020048013

[5] Mori H, Kobara H, Tsushimi T, Nishiyama N, Fujihara S, Masaki T. Unavoidable Human Errors of Tumor Size Measurement during Specimen Attachment after Endoscopic Resection: A Clinical Prospective Study. PloS One 2014; 10(4): e0121798-e.

[6] Khan SK, Shirley MD, Glennie C, Fearon PV, Deehan DJ. Achieving best practice tariff may not reflect improved survival after hip fracture treatment. Clinical Interventions in Aging 2014; 9: 2097. http://dx.doi.org/10.2147/CIA.S65736

[7] Bewick V, Cheek L, Ball J. Statistics review 10: further nonparametric methods. Critical Care-London- 2004; 8(3): 196-200. http://dx.doi.org/10.1186/cc2857

[8] Mirza WA, Imam SH, Kharal M, Aslam M, Ali SA, Masroor I, *et al*. Cleaning methods for ultrasound probes. J Coll Physicians Surg Pak 2008; 18(5): 286-9.

[9] Gamer J, Jarvis W, Emori T, Horan T, Hughes J. CDC definitions for nosocomial infections. Am J Infect Control 1988; 16(3): 128-40.
http://dx.doi.org/10.1016/0196-6553(88)90053-3

[10] Mayon-White R, Ducel G, Kereselidze T, Tikomirov E. An international survey of the prevalence of hospital-acquired infection. Journal of Hospital Infection 1988; 11: 43-8.
http://dx.doi.org/10.1016/0195-6701(88)90164-8

[11] Fowler C, McCracken D. US Probes: Risk of Cross Infection and Ways to Reduce It—Comparison of Cleaning Methods 1. Radiology 1999; 213(1): 299-300.
http://dx.doi.org/10.1148/radiology.213.1.r99au41299

[12] Muradali D, Gold W, Phillips A, Wilson S. Can ultrasound probes and coupling gel be a source of nosocomial infection in patients undergoing sonography? An *in vivo* and *in vitro* study. AJR American Journal of Roentgenology 1995; 164(6): 1521-4.
http://dx.doi.org/10.2214/ajr.164.6.7754907

[13] Kartaginer R, Pupko A, Tepler C. Do Sonographers Practice Proper Infection Control Techniques? Journal of Diagnostic Medical Sonography 1997; 13(6): 282-7.
http://dx.doi.org/10.1177/875647939701300602

[14] Spencer P, Spencer R. Ultrasound scanning of post-operative wounds—the risks of cross-infection. Clinical Radiology 1988; 39(3): 245-6.
http://dx.doi.org/10.1016/S0009-9260(88)80517-8

Received on 16-03-2015 Accepted on 02-05-2015 Published on 21-05-2015

http://dx.doi.org/10.6000/1929-6029.2015.04.02.6