COMPARISON OF POST HOC TEST ON NON-COMMUNICABLE DISEASES DATA

THAMILSELVAN, S.[1*] – AHAD, N. A.[2]

[1] *College of Arts and Sciences, Universiti Utara Malaysia, Kedah, Malaysia.*

[2] *Institute of Strategic Decision Modeling, Universiti Utara Malaysia, Kedah, Malaysia.*

**Corresponding author*
e-mail: swatiselvan[at]gmail.com

(Received 23rd May 2023; accepted 11th August 2023)

Abstract. The Kruskal-Wallis test is a nonparametric statistical test that compares three or more independent sampling groups on a single continuous variable with non-normal distribution. When the Kruskal-Wallis test produces a significant result, it shows that at least one of the samples is different but the test does not show where or how the differences exist. So, post hoc tests are performed to find the significant differences among the sampling groups. This study aims to perform three different post hoc tests on the secondary data sets, identify the significant difference, and compare the performance of the post hoc tests. The post hoc tests namely the Nemenyi test, Dunn's test, and Pairwise Mann-Whitney test are used in this study. The secondary data involved kidney disease, breast cancer, and heart failure. When comparing the performance of the post hoc tests, all three post hoc tests yield the same results for kidney disease and breast cancer data whereas the Pairwise Mann-Whitney test yielded different results compared to the other two post hoc tests for heart failure data. Dunn's test is suitable to be used as the post hoc test as it yields a smaller p-value compared to other tests.
Keywords: *Kruskal Wallis test, Post hoc tests, Nemenyi test, Dunn's test, Pairwise Mann-Whitney test*

Introduction

Parametric tests are tests where a normal distribution of values, also known as a "bell-shaped curve", is assumed (Chin and Lee, 2008). Another name for this distribution type is Gaussian distribution (Chin and Lee, 2008). In general, parametric tests are more powerful than nonparametric tests since it requires a smaller sample size (Chin and Lee, 2008). The most common parametric test is the variance analysis, also known as ANOVA. ANOVA is the most popular statistical method used to evaluate hypotheses. It can accommodate more experimental designs and involves a wide range of subjects (St and Wold, 1989). Nonparametric tests are applied when parametric tests are not suitable. Many nonparametric tests rate the measurements and look for anomalies in the distribution (Chin and Lee, 2008). Nonparametric tests are roughly 95% as effective as parametric tests (Chin and Lee, 2008). One of the nonparametric tests that are commonly used is the Kruskal-Wallis test for one-way ANOVA. The Kruskal-Wallis test is a nonparametric statistical test that compares three or more independent sampling groups on a single continuous variable with non-normal distribution. For data that is not regularly distributed, such as ordinal or rank data, the Kruskal-Wallis test is applicable. On the other hand, the one-way analysis of variance (ANOVA), a parametric test, can be used to examine a continuous variable with a normally distributed distribution. The Kruskal-Wallis test is also known as the two-group Mann-Whitney U (Wilcoxon rank) test. Hence, the Kruskal-Wallis test is an extensive form of the Mann-Whitney U test and a nonparametric version of the one-way ANOVA (McKight and Najab, 2010).

The Kruskal-Wallis test is frequently used to find out if three or more groups vary on a single variable that does not follow the ANOVA's assumptions of normality. The group means cannot be compared since the variable of interest does not fulfil the normality assumption. Instead, the ranks are being compared (McKight and Najab, 2010). Some assumptions are required when conducting the Kruskal-Wallis test. The first assumption required that for the different populations, there are no differences between the continuous distributions for the test variable other than their medians. The second assumption is the test variable scores are independent of each other, and the cases reflect random sampling from the populations (Ostertagova et al., 2014). After performing the Kruskal Wallis test on a set of data and it yields significant results, it shows at least one of the samples is different from the others but does not show where the differences are or how many differences exist. So, post hoc tests are conducted to overcome the problem where multiple comparisons can be run simultaneously. A few post hoc tests namely Nemenyi test, Dunn's test, and Pairwise Mann-Whitney test are chosen in this study and the outcome is compared.

Materials and Methods

Secondary data is applied in this study. Data that was previously collected for another purpose is known as secondary data (Sindin, 2018). For this study, the data is obtained from Kaggle Statistics website. There are three post hoc tests that are applied on the data namely Nemenyi test, Dunn's test, and Pairwise Mann-Whitney test. Microsoft Excel and Statistical Package for Social Science (SPSS) are used to analyze the data.

Nemenyi test

Test statistic will be calculated and compared with the critical value to see if there is any difference between two different groups. The test statistic for the Nemenyi test is shown in Eq. (1).

$$q = \frac{\left|\overline{R_i} - \overline{R_j}\right|}{s.e} \qquad \text{Eq. (1)}$$

Where; $\overline{R_i}$ and $\overline{R_j}$ represents the average rank of the samples in a group. The standard error is as follows:

$$s.e = \sqrt{\frac{k(n+1)}{12}} \qquad \text{Eq. (2)}$$

Where; k denotes the number of groups and n denotes the total sample size where the group size is equal. When the size of the groups is different, the following standard error as in Eq. (3) will be used.

$$s.e = \sqrt{\frac{n(n+1)}{24}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)} \quad \text{Eq. (3)}$$

Where; ni and nj are the sizes of group being prepared. The critical value for the Nemenyi test is obtained from the Studentized Range q table. Then, the critical value is compared with the test statistic. If the test statistic is greater than the critical value, then the conclusion where there exists a difference between groups can be made.

Dunn's test

The calculation of test statistics and comparison with the critical value will indicate whether there is a significant difference between the two groups. The test statistic for the Dunn's test is as follows.

$$z = \frac{|\overline{R_i} - \overline{R_j}|}{s.e} \; where \; \overline{R_j} = \frac{R_j}{n_j} \quad \text{Eq. (4)}$$

Where; the standard error is shown as in Eq. (5).

$$s.e = \sqrt{\frac{n(n+1)}{12}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)} \quad \text{Eq. (5)}$$

When the size of the groups is different, Eq. (6) is used to obtain the standard error.

$$s.e = \sqrt{\left(\frac{n(n+1)}{24} - \frac{(\sum f^3 - f)}{12(n-1)}\right)\left(\frac{1}{n_i}+\frac{1}{n_j}\right)} \quad \text{Eq. (6)}$$

The critical value for Dunn's test is obtained from the normal distribution table and is then compared to the test statistic. If the test statistic exceeds the critical value, it can be concluded that there is a significant difference between the groups.

Pairwise Mann-Whitney test

The test statistics for the Pairwise Mann-Whitney test is given in Eq. (7)

$$T = S - \frac{n_1(n_1+1)}{2} \quad \text{Eq. (7)}$$

Where; S denotes the sum of ranks assigned to the sample from population 1 and n1 is the sample size for sample 1. Obtaining the critical value for the Pairwise Mann-Whitney test involves referring to the Quantiles of Mann-Whitney test statistic table and comparing it to the test statistic. If the test statistic is smaller than the critical value, it indicates a significant difference between the groups.

Results and Discussion

The data is obtained from Kaggle Statistics website. There are 3 types of non-communicable diseases data used in this study which are kidney disease, heart failure, and breast cancer. Kolmogorov-Smirnov test were used to test the normality of the data. Then, Kruskal-Wallis test were conducted to see whether any difference between the groups exist. When significant difference is identified, three post hoc tests are conducted on each data. For this analysis, α=0.05 were used for each test.

Kidney disease

The data for kidney disease consists of three groups which are blood pressure, blood urea, and sodium. *Table 1* below shows the normality test results for kidney disease observations. Based on the output, it is shown that blood pressure, blood urea, and sodium is not normally distributed. *Table 2* presents the results of the Kruskal-Wallis test conducted on the observations of kidney disease. The results indicate a significant difference in medians among the groups. Given the observed significant difference, further analysis was carried out using the Nemenyi test, Dunn’s test, and Pairwise Mann-Whitney test to identify paired differences. *Table 3*, *Table 4* and *Table 5* show the Nemenyi test, Dunn’s test and Pairwise Mann-Whitney test results for kidney disease observations, respectively. The results indicate that for all three pairs Blood Pressure and Blood Urea, Blood Pressure and Sodium, and Blood Urea and Sodium, all the p-values are less than 0.05. Thus, all three pairs are significantly different for all three tests.

Table 1. *Normality test result for kidney disease observation.*

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Blood pressure	.164	117	.000	.933	117	.000
Blood urea	.242	115	.000	.793	115	.000
Sodium	.138	100	.000	.891	100	.000

Table 2. *Kruskal-Wallis test result for kidney disease observation.*

Test statistics	Observations
Kruskal-Wallis H	190.902
df	2
Asymp. Sig.	.000

Table 3. *Nemenyi test result for kidney disease observations.*

Group 1	Group 2	R mean	Std. Error	q-stat	p-value	R-crit
Blood pressure	Blood urea	53.82081	8.912257	6.038966	6.03E-05	29.53522
Blood pressure	Sodium	123.9666	9.243203	13.41165	6.07E-14	30.63197
Blood urea	Sodium	177.7874	9.280168	19.15778	6.07E-14	30.75448

Table 4. *Dunn’s test result for kidney disease observations.*

Group 1	Group 2	R mean	Std. Error	z-stat	R-crit	p-value
Blood pressure	Blood urea	53.82081	12.5861	4.276209	24.66831	1.9E-05
Blood pressure	Sodium	123.9666	13.05347	9.496827	25.58434	0
Blood urea	Sodium	177.7874	13.10568	13.5658	25.68665	0

Table 5. *Pairwise Mann-Whitney test result for kidney disease observations.*

Group 1	Group 2	p-value	U-state	Mean
Blood pressure	Blood urea	6.34E-13	3065	18.98807
Blood pressure	Sodium	0	0	60.6741
Blood urea	Sodium	0	1073	79.66217

Heart failure

The data pertaining to heart failure comprises three groups: creatinine phosphokinase, serum creatinine, and serum sodium. *Table 6* below illustrates the results of the normality tests conducted on the observations of heart failure. Based on *Table 6*, Creatinine Phosphokinase, Serum Creatinine, and Serum Sodium are not normally distributed since the p-value for all three groups is not significant. *Table 7* below shows the Kruskal-Wallis test result for heart failure observations. It is shown that the samples are significantly different. *Table 8* and *Table 9* show the Nemenyi test and Dunn test results for heart failure observations, respectively. For both tests, the pair of Creatinine Phosphokinase and Serum Creatinine, and Serum Creatinine and Serum Sodium are significantly different whereas the pair of Creatinine Phosphokinase and Serum Sodium is not significantly different. *Table 10* shows the Pairwise Mann-Whitney test results for heart failure observations. All three pairs Creatinine Phosphokinase and Serum Creatinine, Creatinine Phosphokinase and Serum Sodium, and Serum Creatinine and Serum Sodium are significantly different.

Table 6. *Normality test result for heart failure observation.*

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Creatinine phosphokinase	.382	100	.000	.368	100	.000
Serum creatinine	.267	100	.000	.592	100	.000
Serum sodium	.114	100	.000	.938	100	.000

Table 7. *Kruskal-Wallis test result for heart failure observations.*

Test statistics	Observations
Kruskal-Wallis H	201.827
df	2
Asymp. Sig.	.000

Table 8. *Nemenyi test result for heart failure observations.*

Group 1	Group 2	R mean	Std. Error	q-stat	p-value	R-crit
Creatinine phosphokinase	Serum creatinine	159.375	8.674676	18.37244	6.0E-14	28.74788
Creatinine phosphokinase	Serum sodium	18.75	8.674676	2.161464	0.277759	28.74788
Serum creatinine	Serum sodium	140.625	8.674676	16.21098	6.07E-14	28.74788

Table 9. *Dunn's test result for heart failure observations.*

Group 1	Group 2	R mean	Std. Error	z-stat	R-crit	p-value
Creatinine phosphokinase	Serum creatinine	159.375	12.26311	12.99629	24.03526	0
Creatinine phosphokinase	Serum sodium	18.75	12.26311	1.528975	24.03526	0.126271
Serum creatinine	Serum sodium	140.625	12.26311	11.46732	24.03526	0

Table 10. *Pairwise Mann-Whitney test result for heart failure observations.*

Group 1	Group 2	p-value	U-state	Mean
Creatinine phosphokinase	Serum creatinine	0	0	547.9824
Creatinine phosphokinase	Serum sodium	0.021956	4062.5	413.37
Serum creatinine	Serum sodium	0	0	134.6124

Breast cancer

The data related to breast cancer comprises three groups: leptin, adiponectin, and resistin. *Table 11* exhibits the normality test results for breast cancer observations. It has been observed that leptin, adiponectin, and resistin exhibit non-normal distribution, as evidenced by the p-values of all three groups being lower than the chosen significance level. *Table 12* shows the result for Kruskal-Wallis test for breast cancer observations. Based on the findings presented in *Table 12*, it is evident that the samples exhibit significant differences, as indicated by the p-value being lower than the predetermined significance level. Therefore, the analysis proceeded with the Nemenyi test, Dunn’s test, and Pairwise Mann-Whitney test as shown in *Table 13*, *Table 14* and *Table 15*. All three pairs, namely Leptin and Adiponectin, Leptin and Resistin, and Adiponectin and Resistin, demonstrate significant differences as the p-values for these pairs are less than 0.05.

Table 11. *Normality test result for breast cancer observations.*

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Leptin	.146	115	.000	.880	115	.000
Adiponectin	.184	115	.000	.825	115	.000
Resistin	.175	115	.000	.746	115	.000

Table 12. *Kruskal-Wallis test result for breast cancer observations.*

Test statistics	Observations
Kruskal-Wallis H	78.118
df	2
Asymp. Sig.	.000

Table 13. *Nemenyi test result for breast cancer observations.*

Group 1	Group 2	R mean	Std. Error	q-stat	p-value	R-crit
Leptin	Adiponectin	115.7586	9.340771	12.39283	6.07E-14	30.95531
Leptin	Resistin	75.38793	9.340771	8.070847	3.89E-08	30.95531
Adiponectin	Resistin	40.37069	9.340771	4.321987	0.006415	30.95531

Table 14. *Dunn’s test result for breast cancer observations.*

Group 1	Group 2	R mean	Std. Error	z-stat	R-crit	p-value
Leptin	Adiponectin	115.7586	13.20984	8.763059	25.89081	0
Leptin	Resistin	75.38793	13.20984	5.706952	25.89081	1.15E-08
Adiponectin	Resistin	40.37069	13.20984	3.056107	25.89081	0.002242

Table 5. *Pairwise Mann-Whitney test result for breast cancer observations.*

Group 1	Group 2	p-value	U-state	Mean
Leptin	Adiponectin	0	2322	16.43421
Leptin	Resistin	5.26E-09	3743	11.88911
Adiponectin	Resistin	0.001423	5097	4.545092

Conclusion

In this study, post hoc tests were utilized to identify paired differences among groups. Specifically, the Nemenyi test, Dunn's test, and Pairwise Mann-Whitney test were conducted as post hoc tests following the Kruskal-Wallis test. Data obtained from the Kaggle Statistics website were used for analysis. Comparing the performance of the three post hoc tests, it was found that all three tests yielded consistent results for the kidney disease and breast cancer data. However, for the heart failure data, the Pairwise Mann-Whitney test produced different results compared to the Nemenyi and Dunn's tests. Specifically, for the kidney disease data, all three pairs (Blood Pressure and Blood Urea, Blood Pressure and Sodium, and Blood Urea and Sodium) were found to be significantly different. For the heart failure data, Nemenyi and Dunn's tests identified a significant difference between the pairs Creatinine Phosphokinase and Serum Creatinine, and Serum Creatinine and Serum Sodium. However, the Pairwise Mann-Whitney test showed a significant difference among all three pairs: Creatinine Phosphokinase and Serum Creatinine, Creatinine Phosphokinase and Serum Sodium, and Serum Creatinine and Serum Sodium. For the breast cancer data, all three pairs (Leptin and Adiponectin, Leptin and Resistin, and Adiponectin and Resistin) were found to be significantly different.

Furthermore, a smaller p-value increases the probability of rejecting the null hypothesis. Based on the results obtained, it can be concluded that Dunn's test is more suitable to be used as the post hoc test for the Kruskal-Wallis test, as it yielded smaller p-values compared to the other post hoc tests.

Acknowledgement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of interest

The authors declare that there is no conflict of interest involve in this research study.

REFERENCES

[1] Chin, R., Lee, Y.B. (2008): Analysis of data: Principles and practice of clinical trial medicine. – Foundations of Anesthesia, 2nd ed.; Academic Press: Cambridge, MA, USA 560p.
[2] McKight, P.E., Najab, J. (2010): Kruskal-Wallis Test. – The Corsini Encyclopedia of Psychology 1: 1-10.
[3] Ostertagova, E., Ostertag, O., Kováč, J. (2014): Methodology and application of the Kruskal-Wallis test. – Applied Mechanics and Materials 611: 115-120.
[4] Sindin, P.X. (2018): Secondary data. – In The SAGE Encyclopedia of Communication Research Methods 2p.
[5] St, L., Wold, S. (1989): Analysis of variance (ANOVA). – Chemometrics and Intelligent Laboratory Systems 6(4): 259-272.