

Performance of Two-Samples Pseudo-Median Procedure (Prestasi Prosedur Pseudo-Median Dua Sampel)

NOR AISHAH AHAD*, ABDUL RAHMAN OTHMAN &
SHARIPAH SOAAD SYED YAHAYA

ABSTRACT

This article investigates the performance of two-sample pseudo-median based procedure in testing differences between groups. The procedure is a modification of the one-sample Wilcoxon procedure using the pseudo-median of differences between group values as the central measure of location. The test was conducted on two groups with moderate sample sizes of symmetric and asymmetric distributions. The performance of the procedure was measured in terms of Type I error and power rates computed via Monte Carlo methods. The performance of the procedure was compared against the t -test and Mann-Whitney-Wilcoxon test. The findings from this study revealed that the pseudo-median procedure performed very well in controlling Type I error rates close to the nominal value. The pseudo-median procedure outperformed the Mann-Whitney-Wilcoxon test and is comparable to the t -test in controlling Type I error and maintaining adequate power.

Keywords: Monte Carlo simulation; power; pseudo-median; Type I error

ABSTRAK

Artikel ini mengkaji prestasi prosedur berasaskan pseudo-median dua sampel dalam menguji perbezaan di antara kumpulan. Prosedur ini terhasil melalui pengubahsuaian prosedur Wilcoxon satu sampel menggunakan pseudo-median semua perbezaan nilai antara kumpulan sebagai ukuran memusat lokasi. Ujian ini dilakukan ke atas dua kumpulan dengan saiz sampel sederhana daripada taburan simetri dan tidak simetri. Prestasi prosedur ini diukur berasaskan Ralat Jenis I dan kadar kuasa yang diperolehi melalui kaedah Monte Carlo. Prestasi prosedur ini dibandingkan dengan ujian- t dan ujian Mann-Whitney-Wilcoxon. Dapatan kajian menunjukkan bahawa prosedur pseudo-median mempunyai prestasi yang sangat baik dalam mengawal kadar Ralat Jenis I hampir kepada aras nominal. Prosedur pseudo-median mengatasi ujian Mann-Whitney-Wilcoxon dan setanding dengan ujian- t untuk mengawal ralat jenis I dan mengekalkan kuasa yang mencukupi.

Kata kunci: Kuasa; pseudo-median; Ralat Jenis I; simulasi Monte Carlo

INTRODUCTION

The conventional procedures of comparing two groups such as Student's t -test are usually restricted by the assumptions of normality and homogeneity of variances. However, in the real world these assumptions are not always fulfilled. It is well known that the Student's t -test can become seriously biased when homogeneity of variance assumption is not satisfied, particularly when the design is unbalanced (Maxwell & Delaney 2004). Over the years, many procedures were developed to handle the violation of these assumptions. Nonparametric procedure such as Mann-Whitney-Wilcoxon is a viable alternative that can be used when the distribution is not normal, while Welch's t -test (Welch 1938) is an example of a procedure that was developed to handle the problem of unequal variances. However, each of these methods specializes on a specific violation only. When multiple problems such as non-normality, variance heterogeneity and unbalanced sizes occur simultaneously, the Type I error will inflate (Keselman et al. 1995; Lix & Keselman 1998). Because of this, the development of new methods for comparing

two groups and locating the differences in the one-way independent groups design require serious attention and remains a very active area of study.

In this study, a pseudo-median based procedure was used in testing the equality of central tendency measures. This procedure was the modification of the one-sample nonparametric Wilcoxon procedure using the pseudo-median of differences between group values as the central measure of location. As stated in Hoyland (1965), the pseudo-medians of a distribution F is defined as the median of the distribution of $(X_1 + X_2)/2$ where X_1 and X_2 are independently and identically distributed according to F . Hollander and Wolfe (1999) noted that the pseudo-median of a distribution F is the median of $(Z_1 + Z_2)/2$, where Z_1 and Z_2 are independent, each with the same distribution F .

The pseudo-median is a location parameter and its value need to be estimated. The estimation was done using the Hodges-Lehmann estimator (Hollander & Wolfe 1999). This study investigated the performance of the pseudo-median procedure in terms of controlling Type I error rates and maintaining adequate power. The performance

was then compared with the parametric and nonparametric counterparts i.e. the t -test and Mann-Whitney-Wilcoxon test, respectively.

PSEUDO-MEDIAN PROCEDURE

The pseudo-median procedure is the modification of the one-sample nonparametric Wilcoxon procedure with the incorporation of pseudo-median of differences between group values as the statistic of interest in a two group setting. In this procedure, let $X_1 = (X_{11}, X_{12}, \dots, X_{1n})$ and $X_2 = (X_{21}, X_{22}, \dots, X_{2m})$ be samples from distributions F_1 and F_2 , respectively. Let the differences between the observations from both samples be $D_{ij} = X_{1i} - X_{2j}$, $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, m$. The absolute value of the differences is given by $|D_{ij}|$ and R_{ij} denotes the rank of $|D_{ij}|$. The indicator function is defined as (1):

$$e_{ij} = \begin{cases} 0, & D_{ij} < 0 \\ 0.5, & D_{ij} = 0 \\ 1, & D_{ij} > 0 \end{cases} \quad (1)$$

The Wilcoxon statistic is defined as:

$$W = \sum_{i=1}^n \sum_{j=1}^m R_{ij} e_{ij} \quad (2)$$

The Hodges-Lehmann estimator of the pseudo-median given in (3) was used to estimate the pseudo-median value where $i \neq i'$ and $j \neq j'$.

$$\begin{aligned} \hat{\theta} &= \text{median} \left(\frac{D_{ij} + D_{i'j'}}{2} \right) \\ &= \text{median} \left[\frac{(X_{1i} - X_{2j}) + (X_{1i'} - X_{2j'})}{2} \right], \end{aligned} \quad (3)$$

The modification of the Wilcoxon procedure was performed by adding the pseudo-median value to all observations in the second sample. A bootstrap procedure was employed to test the hypothesis.

DESIGN SPECIFICATIONS

In evaluating the performance of the test procedures, four variables were considered under several different scenarios or conditions. They were (1) sample sizes, (2) type of distribution – symmetric or asymmetric, (3) group variances and (4) pairing of unequal sample sizes and unequal variances. We used equal sample sizes (20, 20) and unequal sample sizes (15, 25) with total sample sizes of the two groups being $N = 40$. We chose this sample sizes because other researchers have found them to be generally sufficient to provide reasonably effective Type I error control (Wilcox 1994).

For symmetric distribution, we chose Beta (0.5, 0.5) with kurtosis of -1.5. Meanwhile, for asymmetric distribution, we chose one Fleishman (1978) transformation of the standard normal distribution with

skewness of 0.5 and kurtosis of -0.5. Both distributions are chosen to represent platykurtic distribution. Normal distribution was used as the basis of comparison. For the study on other type of distributions, refer to Ahad et al. (2009, 2011).

In terms of group variances, the ratios of 1:1, 1:16 and 1:36 were used in this study. These ratios were chosen so that the performance of the test procedures under equal and unequal variances can be investigated. The 1:36 ratio was chosen as it reflects extreme variance heterogeneity under which the efficacies of the tests should be examined (Alexander & Govern 1994; Keselman et al. 2007; Othman et al. 2004).

The pairings of unequal sample sizes with unequal variances might produce different effects on the Type I error and power rates of many test statistics (Cribbie & Keselman 2003). The two types of pairing to be evaluated are positive and negative pairings. Positive and negative pairings commonly resulted in conservative and liberal Type I error, respectively (Keselman et al. 2007; Othman et al. 2004; Syed Yahaya et al. 2006).

This study was based on simulated data. The simulation was carried out using the random-number-generating function in SAS and the simulation program was written in SAS/IML (SAS 2006). In terms of data generation, pseudo-random normal variates were generated by employing the SAS generator RANDGEN (SAS 2006). This involved the straight forward usage of the (RANDGEN(Y, 'NORMAL')) to generate normal variates with means equals to zero and standard deviation equals to one. Data for Beta (0.5, 0.5) distribution was generated using the RANDGEN subroutine with the beta distribution option, (RANDGEN(Y, 'BETA', 0.5, 0.5)).

Fleishman distribution was obtained via the Fleishman (1978) power transformation. This transformation is of the form $Y = a + bZ + cZ^2 + dZ^3$ where Z is a standard normal variate. Fleishman provided a table of values for the coefficients, b , c , and d in Fleishman (1978) which enables one to transform the standard normal distribution into a non-normal distributions, also of mean zero and variance one. The extra coefficient a is easily obtained through the relation $a = -c$, that was a direct result of constraining $E(Y) = 0$. A set of coefficients (b , c , d) which is (1.1478, 0.1202, -0.0575) were picked from Fleishman (1978) to transform the Z values to produce Fleishman distribution.

The effect size or the shift parameter used in this study to obtain power rates was computed based on the common language (CL) statistic proposed by McGraw and Wong (1992) and A from Vargha and Delaney (2000). The shift parameter varies from 0.2 to 2.0 in increments of 0.2. For each condition examined, 5000 data sets were simulated. The nominal level of significance was set at $\alpha = 0.05$.

MONTE CARLO ALGORITHM

The Monte-Carlo method for getting the Type I error rates for all procedures is given as Algorithm 1. Bootstrap

p -value of the pseudo-median procedure is given as Algorithm 2.

Algorithm 1

1. Initialize a variable, count = 0.
2. Generate two samples data (X_1 and X_2) based on design specification to reflect the null hypothesis.
3. Perform the hypothesis test based on the generated data at the predetermined significance level ($\alpha = 0.05$).
4. Compute and collect the corresponding p -values.
5. If p -value ≤ 0.05 , then increase count by one (count = count + 1).
6. Repeat Step 2 to Step 5 for 5000 times.
7. Calculate the average Type I error rates by dividing count by 5000.

Algorithm 2

1. Based on the two samples, find W statistic and estimate the pseudo-median, ($\hat{\theta}$).
2. Add $\hat{\theta}$ to all observations in the second sample.
3. Calculate \hat{W} from X_1 and new sample, ($X_2 + \hat{\theta}$).
4. Generate bootstrap samples from X_1 and ($X_2 + \hat{\theta}$) yielding X_1^* and X_2^* .
5. Calculate W^* from the bootstrap samples.
6. Calculate ($W^* - \hat{W}$)
7. Repeat step 4 and step 6 for B times. B = 599 appears sufficient in most situations when $n \geq 12$ (Wilcox 2005).
8. Compare the value of ($W^* - \hat{W}$) with ($W - E(W|H_0)$) where $E(W|H_0) = \left[\frac{N(N+1)}{4} \right]$. Let $U = (W^* - \hat{W}) > (W - E(W|H_0))$ and $L = (W^* - \hat{W}) < (W - E(W|H_0))$.

$$9. \quad U = \begin{cases} 1 & \text{if } (W^* - \hat{W}) > (W - E(W|H_0)) \\ 0 & \text{otherwise} \end{cases} \text{ and}$$

$$L = \begin{cases} 1 & \text{if } (W^* - \hat{W}) < (W - E(W|H_0)) \\ 0 & \text{otherwise} \end{cases}$$

10. Calculate the number of times $U = L = 1$.
11. Calculate the p -value as minimum (number of L , number of U) $\times \frac{2}{B}$.

To obtain the power rate for each procedure, the first step is to determine the effect size (0.2, 0.4, 0.6, ..., 2.0) and induced the effect size to all members of the first sample, X_1 . Then, continue all the steps as listed in Algorithm 1 and Algorithm 2.

RESULTS AND DISCUSSION

This study used the Bradley's (1978) liberal criterion of robustness to assess the performance of a statistical test to control its probability of Type I error. According to Bradley's liberal criterion of robustness, a test can be considered robust if its empirical rate of Type I error is within the interval $[0.5\alpha, 1.5\alpha]$. Thus, when the nominal level is set at $\alpha = 0.05$, the procedure or test is considered robust if its' Type I error rate is between 0.025 and 0.075. Type I error rates greater than 0.075 are considered liberal and those less than 0.025 are considered conservative.

The simulation results of Type I error rates under various combinations of group sizes, group variances and types of distribution are presented in Table 1. The results from Table 1 show that all Type I error rates are robust and within the Bradley's liberal criterion of robustness for pseudo-median procedure and t -test. Type I error

TABLE 1. Type I error rates for all procedures under various combinations of study conditions

Procedure	Distribution	(20,20)			(15,25)				
		(1:1)	(1:16)	(1:36)	(1:1)	(1:16)	(16:1)	(1:36)	(36:1)
						+ve	-ve	+ve	-ve
PM	Normal	0.0552	0.053	0.049	0.0526	0.0468	0.0508	0.0486	0.0492
	Beta	0.046	0.0376	0.0348	0.0528	0.0394	0.0402	0.0358	0.0376
	Fleishman	0.0456	0.046	0.0462	0.0528	0.047	0.0572	0.0462	0.0486
t -test	Normal	0.054	0.0532	0.052	0.054	0.0486	0.0508	0.0492	0.0514
	Beta	0.0536	0.0544	0.0544	0.0562	0.051	0.0582	0.0524	0.06
	Fleishman	0.052	0.0536	0.054	0.0532	0.0498	0.056	0.0502	0.0574
MWW	Normal	0.0516	0.079	0.0912	0.0456	0.0408	0.1142	0.0458	0.1142
	Beta	0.0546	0.087	0.0918	0.0492	0.0446	0.1326	0.0474	0.141
	Fleishman	0.0508	0.0976	0.1082	0.0458	0.0594	0.127	0.0614	0.139

Bold values indicate Type I error within [0.025, 0.075]

rates of Mann-Whitney-Wilcoxon test are liberal under conditions of equal sample sizes with heterogeneous variances and unequal sample sizes with negative pairing. All procedures resulted in greater Type I error rates for negative pairing compared with positive pairing regardless of the distributions.

There are no formal standards for power. In determining desired levels of power, most researchers assess the power of their tests using 0.80 as a standard for adequacy. According to Murphy and Myers (2004), power of 0.80 or above is usually judged to be adequate. Therefore, in this study, 0.80 was used as the benchmark for standard adequacy or the desired level of power to be achieved. The results for power analysis were reported only for conditions which have robust Type I error rates.

Tables 2 and 3 display the power rates for all procedures with equal sample sizes and unequal sample sizes, respectively. As can be observed from all distributions, the pseudo-median procedure and the *t*-test are capable of

reaching the 0.80 benchmark only when variances were homogeneous between shift parameter values of 0.8 and 1.0. The power rates for all procedures are vulnerable to variance heterogeneity. Tables 2 and 3 reveal that for both sample sizes, pairing them with heterogeneous variances resulted in depressed statistical power rates. From Table 3, we can also observe that positive pairing led to greater power rates compared with negative pairing for all procedures and distributions.

CONCLUSION

When comparing the differences between two groups, instead of using the classical *t*-test (either Student's *t*-test or Welch's *t*-test), one can use pseudo-median procedure since this procedure can be applied for both variance conditions i.e. equal or unequal. The findings from this study revealed that the pseudo-median procedure performs very well in controlling Type I error rates close to the nominal value. The pseudo-median procedure outperforms

TABLE 2. Power rates for all procedures with equal sample sizes (20,20)

Procedure	Variances	PM			<i>t</i> -test			MWW			
		(1:1)	(1:16)	(1:36)	(1:1)	(1:16)	(1:36)	(1:1)	(1:16)	(1:36)	
Normal	Shift Parameter	0.2	0.095	0.0566	0.051	0.099	0.0584	0.056	0.091	Not reported	Not reported
		0.4	0.231	0.0674	0.054	0.236	0.0702	0.058	0.223		
		0.6	0.450	0.0968	0.069	0.460	0.1006	0.075	0.442		
		0.8	0.681	0.1314	0.089	0.696	0.1428	0.091	0.672		
		1.0	0.860	0.1680	0.104	0.866	0.1768	0.109	0.845		
		1.2	0.956	0.2276	0.126	0.962	0.2404	0.135	0.954		
		1.4	0.990	0.2870	0.149	0.991	0.3004	0.159	0.987		
		1.6	0.998	0.3792	0.199	0.998	0.3960	0.213	0.997		
		1.8	1.000	0.4534	0.235	1.000	0.4682	0.246	1.000		
2.0	1.000	0.5182	0.267	1.000	0.5412	0.282	1.000				
Beta	Shift Parameter	0.2	0.085	0.0420	0.037	0.097	0.0570	0.055	0.124	Not reported	Not reported
		0.4	0.195	0.0532	0.045	0.214	0.0686	0.061	0.263		
		0.6	0.406	0.0732	0.054	0.445	0.0916	0.072	0.467		
		0.8	0.649	0.1130	0.075	0.699	0.1376	0.093	0.666		
		1.0	0.835	0.1436	0.082	0.870	0.1724	0.099	0.804		
		1.2	0.959	0.1918	0.108	0.968	0.2208	0.128	0.913		
		1.4	0.991	0.2666	0.144	0.993	0.2916	0.161	0.960		
		1.6	0.998	0.3240	0.161	0.998	0.3530	0.181	0.986		
		1.8	1.000	0.4022	0.203	1.000	0.4390	0.219	0.998		
2.0	1.000	0.4800	0.245	1.000	0.5346	0.263	0.999				
Fleishman	Shift Parameter	0.2	0.091	0.0532	0.049	0.093	0.0640	0.060	0.093	Not reported	Not reported
		0.4	0.219	0.0576	0.048	0.236	0.0886	0.074	0.228		
		0.6	0.436	0.0822	0.061	0.462	0.1220	0.090	0.446		
		0.8	0.673	0.1072	0.066	0.698	0.1608	0.108	0.669		
		1.0	0.854	0.1492	0.095	0.869	0.1988	0.128	0.838		
		1.2	0.954	0.1932	0.109	0.963	0.2608	0.155	0.949		
		1.4	0.989	0.2352	0.126	0.992	0.3178	0.177	0.984		
		1.6	0.999	0.3016	0.160	0.997	0.4022	0.230	0.995		
		1.8	0.999	0.3668	0.189	1.000	0.4676	0.262	1.000		
2.0	1.000	0.4420	0.224	1.000	0.5316	0.299	1.000				

Bold values indicate power rate ≥ 0.80

TABLE 3. Power rates for all procedures with unequal sample sizes (15,25)

Procedure	PM				t-test				MWW			
	(1:1)	(1:16)	(16:1)	(36:1)	(1:1)	(1:16)	(16:1)	(36:1)	(1:1)	(1:16)	(16:1)	(36:1)
Variances		0.089	0.052	0.053	0.050	0.088	0.054	0.053	0.051	0.082	0.046	0.048
		0.216	0.077	0.057	0.049	0.224	0.080	0.059	0.067	0.207	0.063	0.058
		0.429	0.112	0.083	0.063	0.444	0.113	0.087	0.080	0.411	0.081	0.063
		0.670	0.152	0.112	0.079	0.676	0.160	0.113	0.103	0.647	0.121	0.079
		0.829	0.197	0.146	0.093	0.841	0.214	0.144	0.122	0.812	0.142	0.087
		0.947	0.265	0.177	0.103	0.953	0.280	0.179	0.158	0.937	0.205	0.118
		0.984	0.351	0.232	0.120	0.987	0.365	0.243	0.198	0.980	0.274	0.152
		0.997	0.438	0.286	0.155	0.997	0.457	0.299	0.254	0.996	0.346	0.184
		0.999	0.520	0.350	0.182	0.999	0.545	0.359	0.301	0.999	0.413	0.215
		1.000	0.604	0.409	0.208	1.000	0.629	0.427	0.339	1.000	0.497	0.249
Pairing		0.090	0.046	0.044	0.039	0.098	0.058	0.067	0.054	0.121	0.048	0.049
		0.176	0.058	0.039	0.032	0.191	0.073	0.058	0.060	0.223	0.049	0.050
		0.389	0.084	0.057	0.041	0.432	0.099	0.079	0.076	0.438	0.058	0.056
		0.610	0.142	0.073	0.047	0.663	0.158	0.099	0.103	0.617	0.078	0.063
		0.806	0.178	0.104	0.063	0.845	0.199	0.135	0.116	0.771	0.084	0.064
		0.941	0.255	0.147	0.080	0.954	0.282	0.179	0.156	0.892	0.103	0.070
		0.988	0.331	0.187	0.102	0.990	0.358	0.228	0.191	0.946	0.124	0.082
		0.998	0.389	0.235	0.123	0.999	0.426	0.274	0.218	0.980	0.141	0.086
		0.999	0.491	0.280	0.134	0.999	0.530	0.329	0.274	0.998	0.184	0.109
		1.000	0.556	0.345	0.163	1.000	0.620	0.391	0.326	0.999	0.217	0.120
Beta		0.080	0.050	0.051	0.049	0.090	0.063	0.053	0.060	0.083	0.075	0.073
		0.202	0.060	0.057	0.048	0.223	0.089	0.050	0.076	0.205	0.107	0.098
		0.422	0.097	0.073	0.054	0.432	0.130	0.065	0.098	0.408	0.143	0.114
		0.653	0.122	0.093	0.063	0.675	0.176	0.092	0.119	0.645	0.191	0.143
		0.848	0.179	0.129	0.082	0.842	0.225	0.117	0.134	0.816	0.219	0.152
		0.956	0.221	0.178	0.103	0.954	0.292	0.149	0.175	0.940	0.278	0.183
		0.988	0.280	0.214	0.122	0.988	0.374	0.196	0.215	0.982	0.344	0.229
		0.999	0.362	0.284	0.144	0.998	0.457	0.263	0.268	0.995	0.415	0.254
		0.999	0.441	0.343	0.166	1.000	0.538	0.325	0.311	0.999	0.469	0.290
		1.000	0.529	0.421	0.208	1.000	0.619	0.395	0.353	0.999	0.527	0.317
Fleishman		0.080	0.050	0.051	0.049	0.090	0.063	0.053	0.060	0.083	0.075	0.073
		0.202	0.060	0.057	0.048	0.223	0.089	0.050	0.076	0.205	0.107	0.098
		0.422	0.097	0.073	0.054	0.432	0.130	0.065	0.098	0.408	0.143	0.114
		0.653	0.122	0.093	0.063	0.675	0.176	0.092	0.119	0.645	0.191	0.143
		0.848	0.179	0.129	0.082	0.842	0.225	0.117	0.134	0.816	0.219	0.152
		0.956	0.221	0.178	0.103	0.954	0.292	0.149	0.175	0.940	0.278	0.183
		0.988	0.280	0.214	0.122	0.988	0.374	0.196	0.215	0.982	0.344	0.229
		0.999	0.362	0.284	0.144	0.998	0.457	0.263	0.268	0.995	0.415	0.254
		0.999	0.441	0.343	0.166	1.000	0.538	0.325	0.311	0.999	0.469	0.290
		1.000	0.529	0.421	0.208	1.000	0.619	0.395	0.353	0.999	0.527	0.317

Bold values indicate power rate ≥ 0.80

the Mann-Whitney-Wilcoxon test and was comparable to the *t*-test in controlling Type I error and maintaining adequate power.

ACKNOWLEDGEMENTS

We greatly appreciate the helpful comments of the anonymous referees and editor. Their comments have contributed in the improvement of this article. The work that led to the publication of this paper was funded by the Postgraduate Research Grant Scheme of Universiti Sains Malaysia and supported by the Universiti Utara Malaysia.

REFERENCES

- Ahad, N.A., Othman, A.R. & Syed Yahaya, S.S. 2011. Comparative performance of Pseudo-Median Procedure, Welch's Test and Mann-Whitney-Wilcoxon at Specific Pairing. *Journal of Modern Applied Science* 5(5): 131-139.
- Ahad, N.A., Othman, A.R., Syed Yahaya, S.S. & Padmanabhan, A. R. 2009. New Monte Carlo procedure with pseudo-medians for symmetric non-normal distributions. In Mohd Tahir Ismail and Adli Mustafa (eds.) *5th Asian Mathematical Conference Proceedings (volume III)*, June 2009: 159-165.
- Alexander, R.A. & Govern, D.M. 1994. A new and simpler approximation for ANOVA under variance heterogeneity. *Journal of Educational Statistics* 19: 91-101.
- Bradley, J.V. 1978. Robustness? *British Journal of Mathematical and Statistical Psychology* 31: 144-151.
- Cribbie, R.A. & Keselman, H.J. 2003. Pairwise multiple comparisons: A model comparison approach versus stepwise procedures. *British Journal of Mathematical and Statistical Psychology* 56: 167-182.
- Fleishman, A.I. 1978. A method for simulating non-normal distributions. *Psychometrika* 43: 521-532.
- Hollander, M. & Wolfe, D.A. 1999. *Nonparametric Statistical Methods*. 2nd ed. New York: John Wiley & Sons.
- Hoyland, A. 1965. Robustness of the Hodges-Lehmann estimates for shift. *The Annals of Mathematical Statistics* 36: 174-197.
- Keselman, H.J., Carriere, K.C. & Lix, L.M. 1995. Robust and powerful nonorthogonal analyses. *Psychometrika* 60: 395-418.
- Keselman, H.J., Wilcox, R.R., Lix, L. M., Algina, J. & Fradette, K. 2007. Adaptive robust estimation and testing. *British Journal of Mathematical and Statistical Psychology* 60: 267-293.
- Lix, L.M. & Keselman, H.J. 1998. To trim or not to trim: Tests of location equality under heteroscedasticity and non-normality. *Educational and Psychological Measurement* 58: 409-429.
- Maxwell, S.E. & Delaney, H.D. 2004. *Designing Experiments and Analyzing Data* (2nd ed.) Mahwah, NJ: Erlbaum.
- McGraw, K.O. & Wong, S.P. 1992. A common language effect size statistic. *Psychological Bulletin* 111: 361-365.
- Murphy, K.R. & Myors, B. 2004. *Statistical Power Analysis: A Simple and General Model for Traditional and Modern Hypothesis Tests*. 2nd ed. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Othman, A.R., Keselman, H.J., Padmanabhan, A.R., Wilcox, R.R. & Fradette, K. 2004. Comparing measures of the "typical" score across treatment groups. *British Journal of Mathematical and Statistical Psychology* 57: 215-234.
- SAS Institute Inc. 2006. *SAS Online Doc*. Cary, NC: SAS Institute Inc.
- Syed Yahaya, S.S., Othman, A.R. & Keselman, H.J. 2006. Comparing the "typical scores" across independent groups based on different criteria for trimming. *Metodoloski zvezki* 3: 49-62.
- Vargha, A. & Delaney, H.D. 2000. A critique and improvement of the CL common language effect size statistics of McGraw and Wong. *Journal of Educational and Behavioral Statistics* 25: 101-132.
- Welch, B.L. 1938. The significance of the difference between two means when the population variances are unequal. *Biometrika* 29: 350-362.
- Wilcox, R.R. 1994. A one-way random effects model for trimmed means. *Psychometrika* 59: 289-306.
- Wilcox, R.R. 2005. *Introduction to Robust Estimation And Hypothesis Testing*. New York: Academic Press.
- Nor Aishah Ahad* & Sharipah Soaad Syed Yahaya
School of Quantitative Sciences
UUM College of Arts and Sciences
Universiti Utara Malaysia
06010 Sintok
Kedah, Malaysia
- Abdul Rahman Othman
Robust Statistics Computational Laboratory
School of Distance Education
Universiti Sains Malaysia
11800 Minden
Penang, Malaysia

*Corresponding author; email: aishah@uum.edu.my

Received: 10 June 2011

Accepted: 16 April 2012