

11-1-2011

# Type I Error Rates of the Two-Sample Pseudo-Median Procedure

Nor Aishah Ahad  
*Universiti Utara Malaysia*

Abdul Rahman Othman  
*Universiti Sains Malaysia*

Sharipah Soaad Syed Yahaya  
*Universiti Utara Malaysia*

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

---

## Recommended Citation

Ahad, Nor Aishah; Othman, Abdul Rahman; and Yahaya, Sharipah Soaad Syed (2011) "Type I Error Rates of the Two-Sample Pseudo-Median Procedure," *Journal of Modern Applied Statistical Methods*: Vol. 10 : Iss. 2 , Article 3.  
DOI: 10.22237/jmasm/1320120120

## Regular Articles

# Type I Error Rates of the Two-Sample Pseudo-Median Procedure

Nor Aishah Ahad  
Universiti Utara Malaysia,  
Kedah, Malaysia

Abdul Rahman Othman  
Universiti Sains Malaysia,  
Penang, Malaysia

Sharipah Soaad Syed Yahaya  
Universiti Utara Malaysia,  
Kedah, Malaysia

---

The performance of the pseudo-median based procedure is examined in terms of controlling Type I error for a two independent groups test. The procedure is a modification of the one-sample Wilcoxon statistic using the pseudo-median of differences between group values as the central measure of location. The proposed procedure was shown to have good control of Type I error rates under the study conditions regardless of distribution type.

Key words: Mann-Whitney-Wilcoxon, pseudo-median,  $t$ -test, type I error.

---

### Introduction

Testing the equality of central tendency parameters between two independent samples by controlling Type I error is a common statistical problem. If an underlying distribution is normally distributed with equal population variances, the most suitable test statistic to use is the Student's  $t$ -test. Student's  $t$ , however, is sensitive to non-normal data and heterogeneity of variances. Under these situations, Welch's approximate test (Welch, 1938) usually offers the best practical solution, but this statistic does not adequately control Type I error probabilities under non-normal distributions.

To surmount the problem of non-normality, researchers typically seek nonparametric test alternatives, such as the Mann-Whitney-Wilcoxon, which is believed to be effective against violations of normality. Although ranking methods are often useful when samples are obtained from heavy-tailed distributions, they are influenced by unequal variances

similar to parametric tests (Pratt, 1964; Zimmerman & Zumbo, 1992). Further, nonparametric methods are more appropriate for non-normal symmetric data. Many attempts have been made to deal with asymmetric distributions. In this study, a method to handle the problem of asymmetric data, as well as heterogeneity of variances, is suggested. The method is known as the pseudo-median procedure, where the pseudo-median of differences between group values are employed as the central measure of location with the one-sample nonparametric Wilcoxon procedure in a two group setting. The pseudo-median of a distribution  $F$  is defined to be the median of the distribution  $(Z_1 + Z_2)/2$ , where  $Z_1$  and  $Z_2$  are all possible differences between two observations from each group.  $Z_1$  and  $Z_2$  are independent and have the same distribution as  $F$  (Hoyland, 1965; Hollander & Wolfe, 1999).

The pseudo-median is a location parameter. The estimation of this parameter is accomplished using the Hodges-Lehmann estimator. According to Hollander and Wolfe (1999), the Hodges-Lehmann estimator ( $\hat{\theta}$ ) is a consistent estimator of the pseudo-median, which in general may differ from the median. However, when  $F$  is symmetric, the median and pseudo-median coincide. The pseudo-median is selected as the central measure of location because it is convenient and the asymptotic properties of the pseudo-median are the same as

---

Nor Aishah Ahad is an Academician in the School of Quantitative Sciences. Email: aishah@uum.edu.my. Abdul Rahman Othman is a Professor in the School of Distance Education. Email: oarahman@usm.my. Sharipah Soaad Syed Yahaya is an Associate Professor in the School of Quantitative Sciences. Email: sharipah@uum.edu.my.

median. In this study, the performance of the pseudo-medians procedure in terms of Type I error was measured via Monte Carlo simulation. Because the sampling distribution of this pseudo-median procedure is intractable, the bootstrap method was used to arrive at the significant values.

Methodology

This study addresses both symmetric and asymmetric distribution and the methods applied to the two types of distributions are very different. Let  $X_1 = (X_{11}, X_{12}, \dots, X_{1n_1})$  and  $X_2 = (X_{21}, X_{22}, \dots, X_{2n_2})$  be samples from distributions  $F_1$  and  $F_2$  respectively. The pseudo-median is defined as:

$$\begin{aligned} \hat{d} &= \text{Median} \left( \frac{D_{ij} + D_{i'j'}}{2} \right) \\ &= \text{Median} \left( \frac{(X_{1i} - X_{2j}) + (X_{1i'} - X_{2j'})}{2} \right) \end{aligned} \quad (1)$$

where  $i \neq i'$  and  $j \neq j'$ . When  $F_1$  and  $F_2$  are symmetric,  $d$  can be defined as the difference between the centers of symmetry. Hence, the hypothesis is given as:

$$\begin{aligned} H_0 : d &= 0 \\ \text{versus} \\ H_1 : d &\neq 0. \end{aligned} \quad (2)$$

Let  $D_{ij} = X_{1i} - X_{2j}$ ,  $i = 1, 2, \dots, n_1$ ,  $j = 1, 2, \dots, n_2$  and  $N = n_1 n_2$ . The statistic is a one-sample Wilcoxon statistic based on the  $ND_{ij}$ 's. Let  $R_{ij}$  denote the rank of  $|D_{ij}|$ . The indicator function and the statistic are expressed as:

$$e_{ij} = \begin{cases} 0, & D_{ij} < 0 \\ 0.5, & D_{ij} = 0 \\ 1, & D_{ij} > 0 \end{cases} \quad (3)$$

and

$$W = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} R_{ij} e_{ij}. \quad (4)$$

The modification of the Wilcoxon procedure is performed by adding the pseudo-median value to the second sample to form a new sample,  $X_2 + \hat{d}$ . The aligned difference, based on the location-aligned samples, becomes:

$$\hat{D}_{ij} = X_{1i} - (X_{2j} + \hat{d}) = D_{ij} - \hat{d}. \quad (5)$$

Let  $\hat{R}_{ij}$  denote the rank of  $|\hat{D}_{ij}|$ . The indicator function and the aligned statistic are expressed as:

$$\hat{e}_{ij} = \begin{cases} 0, & \hat{D}_{ij} < 0 \\ 0.5, & \hat{D}_{ij} = 0 \\ 1, & \hat{D}_{ij} > 0 \end{cases} \quad (6)$$

and

$$\hat{W} = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \hat{R}_{ij} \hat{e}_{ij}. \quad (7)$$

Because the second sample was realigned with the estimate  $d$ , it is necessary to find the pseudo sampling distribution for the estimate  $W$ . Use of a bootstrap procedure is proposed in order to construct the hypothesis test. Separately bootstrap  $n_i$  observations from  $X_1$  group and  $n_j$  observations from  $X_2 + \hat{d}$  group to obtain bootstrap samples,  $X_1^*$  and  $X_2^*$ . The bootstrap difference becomes  $D_{ij}^* = X_{1i}^* - X_{2j}^*$  where  $R_{ij}^*$  denotes the rank of  $|D_{ij}^*|$ . The indicator function and the bootstrap statistic can be defined as:

$$e_{ij}^* = \begin{cases} 0, & D_{ij}^* < 0 \\ 0.5, & D_{ij}^* = 0 \\ 1, & D_{ij}^* > 0 \end{cases} \quad (8)$$

and

$$W^* = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} R_{ij}^* e_{ij}^*. \quad (9)$$

## TYPE I ERROR RATES OF THE TWO-SAMPLE PSEUDO-MEDIAN PROCEDURE

The steps to obtain the  $p$  value using the bootstrap method for symmetric distribution are as follows:

1. Calculate  $W$  from  $X_1$  and  $X_2$ .
2. Calculate  $\hat{d}$  from  $X_1$  and  $X_2$ .
3. Add  $\hat{d}$  to  $X_2$  to form a new sample,  $X_2 + \hat{d}$ .
4. Calculate  $\hat{W}$  from  $X_1$  and the new sample in step 3.
5. Generate bootstrap samples by randomly sampling with replacement  $n_i$  observations from the  $X_1$  group, and  $n_j$  observations from the new sample in step 3 yielding  $X_1^*$  and  $X_2^*$ .
6. Calculate  $W^*$  from the bootstrap samples,  $X_1^*$  and  $X_2^*$ .
7. Find  $(W^* - \hat{W})$ .
8. Repeat Steps 5 - 7  $B$  times.
9. Compare the value of  $(W^* - \hat{W})$  with  $(W - E(W | H_0))$ .

$$\text{Let } U = (W^* - \hat{W}) > (W - E(W | H_0)) \text{ and} \\ L = (W^* - \hat{W}) < (W - E(W | H_0)).$$

10. Calculate the  $p$  value as  $\frac{2}{B} \times \min(\#L, \#U)$ .

For asymmetric distributions, the difference between the centers of symmetry between the two groups cannot be assumed to be zero; therefore, to ensure the setting for the null condition, a constant  $a$  must be determined and added to the members of the second sample. The value of  $a$  is obtained via simulation. For

example, let  $X_1$  and  $X_2$  be two skewed distributions where the standard deviations need not be the same. Let  $Y_1 = (Y_{11}, Y_{12})$  and  $Y_2 = (Y_{21}, Y_{22})$  represent the new generated samples of size two, which have the same distribution with  $X_1$  and  $X_2$ , respectively. Compute  $a_i$  as follows:

$$a_i = \text{median} \left[ \frac{(Y_{11} - Y_{21}) + (Y_{12} - Y_{22})}{2} \right] \quad (10)$$

Repeat the process of generating new samples of size two 9,999 times and repeat the computation of  $a_i$  to obtain  $a_1, a_2, \dots, a_{10,000}$ . Therefore, the median of  $a_1, a_2, \dots, a_{10,000}$  is the value of  $a$ .

For asymmetric distributions, the steps to obtain the  $p$  value using a bootstrap method are the same except for one small alteration in step 1. In this step, a constant  $a$  is introduced to the members of the second sample ( $X_2$ ) to form a new sample,  $X_{2,new}$ . Steps 2-10 proceed as noted, with the one difference that  $X_2$  has become  $X_{2,new}$ .

To study the robustness of this procedure, four variables were manipulated to create conditions known to highlight the strengths and weaknesses of the test for the equality of location parameters. The variables are (1) types of distributions, (2) degree of variance inequality, (3) balanced/unbalanced sample sizes, and (4) pairings of unequal group variance and sample sizes. In this study, empirical Type 1 error rates were collected and later compared under various study conditions.

The number of groups and sample sizes were fixed. This study covered only the two groups case with total sample size of  $N = 40$ . This value was later divided into two groups forming the balanced and unbalanced design. For the balanced design, the value is equally divided into  $n_1 = n_2 = 20$ , and for the unbalanced design the groups were divided into  $n_1 = 15$  and  $n_2 = 25$ . To investigate the distribution types, this study focused on (1) heavy tailed symmetric non-normal distribution, and (2) heavy tailed asymmetric distribution.

The normal distribution was used as the basis for comparison. The symmetric non-normal distribution was generated from a *g*-and-*h* distribution (Hoaglin, 1985); specifically,  $g = 0$  and  $h = 0.225$  with skewness ( $\gamma_1$ ) = 0 and kurtosis ( $\gamma_2$ ) = 154.84 was chosen for investigation. The Chi-square with three degrees of freedom ( $\gamma_1 = 1.63$  and  $\gamma_2 = 4$ ) was selected to represent the asymmetric distribution.

The pseudo-random normal variates were generated using the SAS generator RANDGEN function (SAS Institute, 1999); this involved the (RANDGEN(Y, 'NORMAL')) function to generate normal variates with means equals to zero and standard deviation equals to one. To generate data from the *g*-and-*h* distribution, standard unit normal variables ( $Z_{ij}$ ) were converted to the *h* random variates via

$$Y_{ij} = Z_{ij} \exp\left(\frac{hZ_{ij}^2}{2}\right). \quad (11)$$

For the Chi-square distribution, data were generated using the (RANDGEN(Y, 'CHISQUARE', 3)) function.

Apart from the types of distribution, two other manipulated variables were the degrees of variance inequality and pairings of variances and group sizes. The nature of pairings of variances and sample sizes affect Type I error rates (Keselman, et al., 1998; Keselman, Othman, Wilcox & Fradette, 2004; Othman, et al., 2004). The variances were manipulated in the following manner: In the case of equal variances, both group variances were set at 1; for the unequal case, the variances were set at 1 and 36.

For positive pairings, the group with the largest number of observations was paired with the group having largest variance, and the group with the smallest number of observations was paired with the group having smallest variance. For the negative pairings, the group with largest number of observations was paired with the group having the smallest variance, and the group with smallest number of observations was paired with the group having the largest variance. This condition was included in the investigation because the direction

(positive/negative) of the pairings has been shown to exert some effect on the results. Positive pairings typically produce conservative results and negative pairings tend to produce liberal results (Keselman, Wilcox, Othman & Fradette, 2002; Cribbie & Keselman, 2003; Othman, et al., 2004; Syed Yahaya, Othman & Keselman, 2004, 2006). Therefore, both positive and negative pairings were evaluated.

The operating characteristics of the procedures investigated in this study could be described as extreme because they substantially depart from homogeneity and normality. These conditions were used because it is reasonable to assume that, if a procedure works under the most extreme conditions, it will probably also work under most conditions likely to be encountered by researchers.

The simulation program was written in SAS/IML (SAS Institute, 1999). For each condition examined, 5,000 data sets were generated and within each data set, 599 bootstrap samples were obtained. The level of significance was set at  $\alpha = 0.05$ .

## Results

To evaluate whether the test is robust (insensitive to assumption violations) under each particular condition, the Bradley criterion of robustness (Bradley, 1978) was employed. According to this criterion, for the five percent nominal level used in this study, a test is considered robust if its empirical Type I error rate is within [0.025, 0.075]. Correspondingly, a test is considered to be non-robust if, for a particular condition, its Type I error rate is not contained in this criterion. This criterion was chosen because it provides a reasonable standard for judging robustness. The empirical Type I error rates for the pseudo-median procedure (PM), *t*-test and Mann-Whitney-Wilcoxon (MWW) across all distributions are displayed in Table 1.

With respect to the procedures, results show that all Type I error rates for the pseudo-median procedure are robust under Bradley's liberal criterion and are very close to the nominal level (0.05) regardless of distribution or conditions. The disparity between Type I error rates from balanced and unbalanced designs is minuscule and the rates are consistent across the

## TYPE I ERROR RATES OF THE TWO-SAMPLE PSEUDO-MEDIAN PROCEDURE

investigated conditions. The  $t$ -test also produces robust Type I error rates for all distributions and conditions, however, for the Chi square distribution, the Type I error rates inflate to a level above 0.065 when the variances are unequal and worsen under negative pairing. For the Mann-Whitney-Wilcoxon test, half of the Type I error rates are above the robustness criterion under unequal variances, especially negative pairing. The Type I error rates for MWW under the Chi-square distribution are too liberal and not robust except under the homogeneous variance condition.

In terms of distributional shapes, the Chi-square distribution produced better empirical Type I error rates compared to the  $g$ -and- $h$  distribution in most conditions for the pseudo-median procedure. Higher values of Type I error rates from Chi-square distribution are apparent for the  $t$ -test and Mann-Whitney-Wilcoxon.

With respect to variance equality and inequality, results show a contradiction between symmetric and asymmetric distributions for both the pseudo-median and the  $t$ -test. For the  $g = 0, h = 0.225$  distributions, homogeneous variances produced greater Type I error rates compared to heterogeneous variances. For the Chi-square distribution, homogeneous variances produced smaller Type I error rates compared to heterogeneous variances. However, no specific pattern could be identified for the Mann-Whitney-Wilcoxon test.

With respect to the pairings of group sizes and variances, results show that the  $g$ -and- $h$  distribution produced liberal ( $> 0.05$ ) Type I error rates for the pseudo-median procedure and conservative ( $< 0.05$ ) results for the  $t$ -test. The Chi-square distribution for the pseudo-median procedure produced conservative Type I error rates for the positive pairing, and liberal results for the negative pairing. The  $t$ -test produced liberal results for both pairings

Table1: Empirical Type I Error Rates of Pseudo-Medians Procedure,  $t$ -test and Mann-Whitney- Wilcoxon\*

Method	Distribution	Group Sizes			
		(20, 20)		(15, 25)	
		Variance (1:1)	Variance (1:36)	Variance (1:36) +ve pairing	Variance (36:1) -ve pairing
PM	Normal $g=0, h=0.225$ $\chi^2_3$	0.0552	0.049	0.0486	0.0492
		0.0588	0.0544	0.0518	0.0532
		0.0454	0.0504	0.0476	0.055
$t$ -test	Normal $g=0, h=0.225$ $\chi^2_3$	0.054	0.052	0.0492	0.0514
		0.0522	0.0458	0.0448	0.044
		0.052	0.0696	0.0654	0.0736
MWW	Normal $g=0, h=0.225$ $\chi^2_3$	0.0516	<b>0.0912</b>	0.0458	<b>0.1142</b>
		0.0516	<b>0.0854</b>	0.0436	<b>0.108</b>
		0.052	<b>0.2428</b>	<b>0.1812</b>	<b>0.2398</b>

\*Bolted entries indicate Type I error rates of the test exceeding the 0.075 criterion.

## Conclusion

The purpose of this study was to investigate how well the pseudo-medians procedure responded to the violations of assumptions compared to the traditional  $t$ -test and Mann-Whitney-Wilcoxon method. The procedure was tested the heavy-tailed distributions, namely the  $g = 0$  and  $h = 0.225$  and the Chi-square with three degrees of freedom. Results show that the Type I error rates for the pseudo-median procedure and the  $t$ -test are robust under Bradley's criterion of robustness and close to the nominal value. The nature of the sample sizes - balanced or unbalanced - did not show much difference in the procedure's ability to control Type I error rates.

The pseudo-median procedure performed better than  $t$ -test, especially for a skewed distribution with unbalanced design and heterogeneous variances. This procedure also outperforms the popular Mann-Whitney-Wilcoxon method in most conditions. The pseudo-median procedure was observed to have good control of Type I error rates, regardless of distributions under the study conditions. The pseudo-median procedure can thus be recommended as an alternative for testing the differences between two groups, particularly when assumptions of normality and variance homogeneity are not met.

## References

- Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, 31, 321-339.
- Cribbie, R. A., & Keselman, H. J. (2003). The effects of nonnormality on parametric, nonparametric, and model comparison approaches to pairwise comparisons. *Educational and Psychological Measurement*, 63(4), 615-635.
- Hoaglin, D. C. (1985). Summarizing shape numerically: The  $g$ - and  $h$ - distributions. In *Exploring data tables, trends, and shapes*, D. C. Hoaglin, F. Mosteller & J. W. Tukey (Eds.), 461-513. New York: NY: Wiley.
- Hollander, M., & Wolfe, D. A. (1999). *Nonparametric statistical methods* (2<sup>nd</sup> Ed.). New York: NY: Wiley.
- Hoyland, A. (1965). Robustness of the Hodges-Lehmann estimates for shift. *The Annals of Mathematical Statistics*, 36, 174-197.
- Keselman, H. J., et al. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Review of Educational Research*, 68(3), 350-386.
- Keselman, H. J., et al. (2004). The new and improved two-sample  $t$ -test. *American Psychological Society*, 15(1), 57-51.
- Keselman, H. J., et al. (2002). Trimming, transforming statistics, and bootstrapping: Circumventing the biasing effects of heteroscedasticity and nonnormality. *Journal of Modern Applied Statistical Methods*, 1(2), 288-309.
- Othman, A. R., et al. (2004). Comparing measures of the "typical" score across treatment groups. *British Journal of Mathematical and Statistical Psychology*, 57(2), 215-234.
- Pratt, J. W. (1964). Robustness of some procedures for two-sample location problem. *Journal of the American Statistical Association*, 59, 665-680.
- SAS Institute, Inc. (1999). *SAS/IML User's Guide version 8*. Cary, NC: SAS Institute.
- Syed Yahaya, S. S., Othman, A. R., & Keselman, H. J. (2004). Testing the equality of location parameters for skewed distributions using  $S_1$  with high breakdown robust scale estimators. In *Theory and Applications of Recent Robust Methods, Series: Statistics for Industry and Technology*, M. Hubert, G. Pison, A. Struyf & S. Van Aelst (Eds.), 319-328. Birkhauser: Basel, Switzerland.
- Syed Yahaya, S. S., Othman, A. R., & Keselman, H. J. (2006). Comparing the "typical score" across independent groups based on different criteria for trimming. *Methodoloski Zvezki-Advances in Methodology and Statistics*, 3, 49-62
- Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika*, 29, 350-362.
- Zimmerman, D. W., & Zumbo, B. D. (1992). Parametric alternatives to the Student  $t$  test under violation of normality and homogeneity of variance. *Perceptual Motor Skills*, 74, 835-844.