The Robustness of the Modified *H*-Statistic in the Test of Comparing Independent Groups

Suhaida Abdullah*, Teh Kian Wooi, Sharipah Soaad Syed Yahaya and Zahayu Md Yusof

College of Art and Sciences, Univerisiti Utara Malaysia, 06010 Sintok Kedah, Malaysia

The *H*-statistic is a robust test statistic in comparing the equality of two and more than two independent groups. This statistic is one of a good alternative to the *F*-statistic in the analysis of variance (*ANOVA*). The *F*-statistic is good only when the distribution of data is normal with homogeneous variances. If there is a violation of at least one of these assumptions, it affects the Type I error rate of the test. The main weakness of the *F*-statistic is its calculation based on the mean. The mean is well-known as a very sensitive central tendency measure with 0 breakdown point, whereas the *H*-statistic provides a test with fewer assumptions yet powerful. This statistic is readily adaptable to any measure of central tendency, and it appears to give reasonably good results. Hence, this paper provides a detailed study on the robustness of the *H*-statistic and its performance using different robust central tendency measures such that the modified one-step *M* (*MOM*) estimator and Winsorized *MOM* estimator. Based on the simulation study, this paper also investigates the performance of the *H*-statistic under various data conditions. The findings reveal that this statistic performs as well as the *F*-statistic under normal and homogeneous variance, yet it provides better control of Type I error rate under non-normal data or heterogeneous variances or both.

Keywords: *H*-statistic; robust test; mean; modified one-step *M*-estimator

I.   INTRODUCTION

Statistical hypothesis testing provides approaches to generalisation based on sample data. The tests are various depend on the aim of the study. In the case of identifying the difference between independent groups, the *t*-test and analysis of variance (also known as *ANOVA F*-test) are the two classical tests that are well known and widely used. These tests are very powerful if the data distribution is normal with homogeneous variances. These assumptions (normality and homogenous variances) are sometimes hard to attain in areal-life situation. What happened to the testing process if there is a violation of the assumptions? There are many of the previous studies define the answer for this question from which revealed the weaknesses of these methods in terms of controlling the Type I error rates (Erceg-Hun & Mirosevich, 2008; Ramussen, 1989; Siegel, 1957; Abdullah, Syed Yahaya & Md Yusof, 2017).

The nonparametric approach is one of the solutions when dealing with the violation of the assumptions in the tests above. Using the nonparametric, it reduces the effect of data distribution because it uses ranking instead of the original data. However, this approach is not the best option since it has low power and might lose some of the important information (Siegel, 1957).

Hence, some of the studies turned to focus more on robust hypothesis testing to produce a better test statistic that provides fewer assumptions yet powerful in identifying a difference. There are a various number of robust test statistics for the case of comparing independent groups such as the Welch test and James test that cover more on dealing with heterogeneous variances (Welch, 1951). While Othman, Keselman, Padmanabhan, Wilcox and Fradette (2004) are more interested in

studying on the *H*-statistic which more robust under non-normal data distribution.

The *H*-statistic was first promoted by Schrader and Hettmansperger (1980) with its specialty where any central tendency measure is readily adapt. This advantage was taken by Othman *et. al.*, (2004) and Syed Yahaya (2005) to propose the use of modified one step *M* (*MOM*) estimator with different scale estimators. Their innovations to the test statistic improved the ability of the test in controlling the Type I error rates under non-normal data conditions. The *MOM* estimator uses the outlier detection as its trimming criteria, but the outlier is not easy to detect. It may cause result in not detecting any, and the power rate will be affected (Wilcox, 2003). Trimming the outlier will eliminate the effect of the outlier but at the same time will reduce the number of the sample. In the case of univariate, this will be not much problem, but in some cases, the reduced sample size could create another problem, especially when it involves the influence of any dependent variable.

Hence, this study aims to produce the modified *H*-statistic that robust to the non-normality and the heterogeneous variances also at the same time preserve the number of the sample size where there will be no outlier will be trimmed out. Therefore, this study aims to promote the use of Winsorization process in the *H*-statistic. The robustness of the proposed Winsorized H-statistic (*WMOM-H*) was evaluated based on its ability to control the Type I error rates. A comparison study between the existing *H*-statistic also being considered to identify the performance of the proposed tests.

II.   MATERIALS AND METHODS

In this study, the Winsorized *MOM* estimator in *H*-statistic (*WMOM-H*) is proposed to eliminate the effect of non-normality while preserving the original sample size. The Type I error rates were calculated to evaluate the performance of the proposed method using a simulation study.

This study manipulated variables such as the distribution, number of groups (*J*), sample sizes, variances and the nature of pairing to create conditions which highlight the strength and weakness of the test. Table 1 presents the design specification of these manipulation variables.

Table 1: Design specification for manipulated variables

J	Sample Size				Variance				Pairing	Label of the condition
	20	20			1	1				C1
	20	20			1	36				C2
2	15	25			1	1				C3
	15	25			1	36			+	C4
	15	25			36	1			−	C5
	20	20	20	20	1	1	1	1		C6
	20	20	20	20	1	1	1	36		C7
4	10	15	25	30	1	1	1	1		C8
	10	15	25	30	1	1	1	36	+	C9
	10	15	25	30	36	1	1	1	−	C10

Based on Table 1, there are three types of distributions considered in this study, which are all represented by the *g-h* distribution. The *g* controls the skewness of the data while the *h* controls the kurtosis. The distribution indicates that for *g=h=0* data is having a normal distribution, whereas, for *g=0:h=0.5, g=1:h=0* and *g=1:h=0.5* are presenting heavy-tailed, skewed normal tailed and skewed heavy-tailed respectively.

The number of two and four groups were chosen to represent conditions of a small and moderate number of groups. The study also controls the sample sizes in the groups so that it is not exceeding 30 to avoid a large sample size condition according to the central limit theorem. In the sample sizes, the balanced and unbalanced sample size also being considered because it will also affect the ability of the test in controlling the Type I error rate (Wilcox, 2003).

The variances homogeneity or equal variance is another assumption required to follow when conducting the traditional statistical test. Two ratios of variances were selected to investigate the performance of the proposed method under equal and unequal variances that are 1:1 and 1:36, respectively. For the unequal variances, the variance ratio chosen is 1:36 according to previous studies which it seems large and reasonable to evaluate the proposed methods' performance under a 'potentially' extreme condition (Keselman *et. al.,* 2007; Syed Yahaya, 2005; Wilcox, 2003;).

The nature of pairing formed when unbalanced sample size paired with unequal variances, and it might provide different results in terms of the Type I error rates (Keselman *et. al.,* 2007; Syed Yahaya, 2005; Wilcox, 2003;). The positive pairing is the cases that the smallest sample size (*n*) paired with the smallest variance, and the largest *n* paired with the largest variance. On the other hand, the negative pairing is the cases that the smallest *n* is paired with the largest variance, and inversely. C1 to C10 label the combination of

the variable number of groups, sample size, variances and the nature of pairing. Therefore, in total, after considering four types of distribution, there are 40 conditions considered in the study.

For simulation study, 5000 simulated datasets were generated using SAS generator RANNOR (SAS Institute Inc., 2011) with 599 bootstrap samples generated to perform the H-statistic. The 5% level of significance (α = 0.05) uses to determine the performances of the methods.

III.   RESULTS AND DISCUSSION

The robustness of the proposed *WMOM-H* evaluated based on its ability to control the Type I error rates in the simulation study. The Type I error was set to be equal 0.05 so that the rates must be around this value. Thus Bradley (1978) provided criteria to the Type I error rates to evaluate the robustness of a test where the test is robust if the Type I error rates are from 0.5α to 1.5α. For α = 0.05, the Type I error rates are within 0.025 to 0.075. Table 2 to Table 5 depict the results of the Type I error rates for all compared tests. The bolded values are the Type I error rates that robust and nearest to the nominal level 0.05.

Table 2 depicts the results of Type I error rates for all compared tests under the normal distribution. The *WMOM-H* and *MOM-H* are robust under all considered conditions with the Type I error rates within 0.025 to 0.075. However, the *MOM-H* are more robust to the heterogeneous variance compared to *WMOM-H,* where the *MOM-H* can have better control of the Type I error rates in more conditions. For two group case (represented by C1 to C5), the t-test is still considered robust if the sample sizes are balanced and variances are homogeneous (C1 to C3). While the *ANOVA F*-test is only robust under homogeneous variances.

Under symmetric heavy-tailed distribution which represented by the g=0:h=0.5 as in Table 3, seems like the *WMOM-H* remain its robustness and has better performance compared to the *MOM-H*. The bolded values represent the nearest value to 0.05 shows that the *WMOM-H* having most conditions with these values. However, it is observable that the *WMOM-H* not very well performs under homogeneous variances, but it still considered robust. For the *MOM-H,* it found to be not robust with conservative Type I error rates under homogeneous variances. While for two group case (C1 to C5), the t-test still robust under balanced sample size or homogeneous variances. The *ANOVA F*-test (C6 to CC10) still able to remain it robustness if the variances are homogeneous.

Table 4 displays the result of Type I error rates under skewed normal tailed distribution. The *WMOM-H* fail to control the Type I error under positive pairing for two group case (C4), and under balanced sample and homogeneous variances for four group case (C8). At the meantime, the *MOM-H* test becomes not robust under the condition of homogeneous variances (C6 and C8). However, this test performs well compared the other test when it has six out of ten conditions with Type I error rates nearest to 0.05. The *ANOVA F*-test still able to control the Type I error rates even though under skewed distribution as long the variances are homogeneous.

Table 5 provides results of the Type I error rates under skewed heavy-tailed distribution and revealed that the proposed test (WMOM-H) consistently robust with the nearest value to 0.05 in all conditions for two group case and four group case, it only not robust when variances are homogeneous.

Table 2. The Type I error rates under normal distribution

Distribution		Label of the condition	WMOM-H	MOM-H	t-test/ F-test
g	h				
		C1	**0.0526**	0.0410	0.0528
		C2	0.0646	**0.0498**	0.0618
		C3	0.0474	0.0366	**0.0490**
		C4	0.0628	**0.0496**	0.0198
0	0	C5	0.0570	**0.0470**	0.1268
		C6	0.0420	0.0256	**0.0518**
		C7	0.0592	**0.0460**	0.1096
		C8	0.0388	0.0246	**0.0504**
		C9	0.0644	**0.0486**	0.0336
		C10	0.0622	**0.0528**	0.2850

Table 3. The Type I error rates under symmetric heavy tailed distribution

Distribution		Label of the condition	WMOM-H	MOM-H	t-test/ F-test
g	h				
		C1	**0.0412**	0.0214	0.0356
		C2	**0.0474**	0.0324	0.0402
		C3	0.0366	0.0232	**0.0374**
		C4	**0.0496**	0.0304	0.0118
0	0.5	C5	**0.0450**	0.0330	0.0996
		C6	0.0248	0.0078	**0.0336**
		C7	**0.0472**	0.0292	0.0782
		C8	0.0194	0.0076	**0.0404**
		C9	**0.0444**	0.0302	0.0192
		C10	**0.0372**	0.0274	0.2392

Table 4. The Type I error rates under skewed normal tailed distribution

Distribution		Label of the condition	WMOM-H	MOM-H	t-test/ F-test
g	h				
		C1	**0.0424**	0.0312	0.0358
		C2	0.0688	**0.0544**	0.1226
		C3	**0.0442**	0.0300	0.0382
		C4	0.0764	**0.0506**	0.0370
		C5	0.0620	**0.0580**	0.2334
1	0	C6	0.0234	0.0134	**0.0432**
		C7	0.0634	**0.0512**	0.2448
		C8	0.0216	0.0140	**0.0442**
		C9	0.0766	**0.0476**	0.1278
		C10	0.0656	**0.0616**	0.3804

Table 5. The Type I error rates under skewed heavy tailed distribu

Distribution		Label of the condition	WMOM-H	MOM-H	t-test/ F-test
g	h				
		C1	**0.0324**	0.0200	0.0232
		C2	**0.0468**	0.0398	0.0434
		C3	**0.0330**	0.0178	0.0272
		C4	**0.0500**	0.0374	0.0088
		C5	**0.0448**	0.0356	0.1108
1	0.5	C6	0.0164	0.0070	0.0226
		C7	**0.0438**	0.0336	0.0918
		C8	0.0138	0.0052	**0.0376**
		C9	**0.0430**	0.0334	0.0278
		C10	**0.0450**	0.0314	0.2214

IV. CONCLUSIONS

After the observations on the ability of the proposed test (the *WMOM-H*), the Winsorization process does improve the ability of the *H*-statistic, especially under heavy-tailed distribution. Even though it does not perform very well under skewed distribution like the *MOM-H*, but the performance of this test still can be considered good with robust under several conditions. Using the Winsorization in the *H*-statistic (the *WMOM-H*) is just like a compliment to the *MOM-H* that handle the problem of data skewed only but not the heavy-tailed distribution. Besides it helps to handle the problem of heavy-tailed distribution, the *WMOM-H* also remains the number of sample size.

V. ACKNOWLEDGEMENTS

VI. REFERENCES

Abdullah, S. Syed Yahaya, S. S & Md Yusof, Z. 2017, "The robust test statistic in comparing two independent groups using trimming and Winsorization", International Journal of Basics and Applied Sciences, vol.17, no.6.

Bradley, J. V. 1978, "Robustness?" in British Journal of Mathematical and Statistical Psychology, 31, pp. 144-152.

Erceg-Hurn, D. M. & Mirosevich, V. M. 2008, "Modern robust statistical methods," in American Psychologist, 2008, pp. 591-601.

Keselman, H. J. Wilcox, R. R. Lix, L. M. Algina, J. & Fradette, K. 2007, "Adaptive robust estimation and testing," in British Journal of Mathematical and Statistical Psychology, 60, pp. 267–293.

Othman, A. R. Keselman, H. J. Padmanabhan, A R. Wilcox, R. R. & Fradette, K. 2004, "Comparing measures of the "typical" score across treatment groups," in British Journal of Mathematical and Statistical Psychology, 57(2), pp. 215-234.

Rasmussen, J. L. 1989, "Data transformation, Type I error rate and power," in British Journal of Mathematical and Statistical Psychology, 42, pp. 203–213.

Reed, J. F. & Stark, D. B. 1996, "Hinge estimators of location: robust to asymmetry," in Computer Methods and Programs in Biomedicine, 49, 1996, pp. 11-17.

SAS Institute Inc. 2011, SAS/IML User's Guide version 9.3., SAS Institute Inc., Cary, NC, 2011.

Schrader, R. M. & Hettmansperger, T. P. 1980, "Robust Analysis of Variance Based Upon a Likelihood Ratio Criterion," in Biometrika, 67, pp. 93-101.

Siegel, S. 1957, "Nonparametric Statistics," in The American Statistician, 11(3), pp. 13-19.

Syed Yahaya, S. S. 2005, "Robust statistical procedures for testing the equality of central tendency parameters under skewed distributions," Unpublished PhD Thesis, Universiti Sains Malaysia.

Welch, B. L. 1951, "On the comparison of several mean values: an alternative approach," in Biometrika, 38, pp. 330-336.

Wilcox, R. R. 2003, Applying contemporary statistical technique, Academic Press, San Diego, CA.