Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.12785/amis/072L04

Comparing the performance of modified F_t statistic with ANOVA and Kruskal Wallis test

Zahayu Md Yusof[1], *Suhaida Abdullah*[2] *and Sharipah Soaad Syed Yahaya*[3]

[123]School of Quantitative Sciences, Universiti Utara Malaysia, 06010 UUM Sintok, Kedah

Received: 8 Oct. 2012, Revised: 4 Jan. 2013, Accepted: 6 Jan. 2013
Published online: 1 Jun 2013

Abstract: ANOVA is a classical test statistics for testing the equality of groups. However this test is very sensitive to nonnormality as well as variance heterogeneity. To overcome the problem of nonnormality, robust method such as F_t test statistic can be used but the test statistic can only perform well when the assumption of homoscedasticity is met. This is due to the biasness of mean as a central tendency measure. This study proposed a robust procedure known as modified F_t method which combines the F_t statistics with one of the popular robust scale estimators, MAD_n, T_n and LMS_n. A simulation study was conducted to compare the robustness (Type I error) of the method with respect to its counterpart from the parametric and non parametric aspects, ANOVA and Kruskal Wallis respectively. This innovation enhances the ability of modified F_t statistic to provide good control of Type I error rates. The findings were in favor of the modified F_t method especially for skewed data. The performance of the method was demonstrated on real education data

Keywords: robust, Type I error, skewed distributions, education

1 Introduction

Classical statistical methods such as ANOVA which are frequently used by researchers to test their work are confined to certain assumptions. One of the assumptions is that the population under study is normally distributed. The uninformed usage of this method under violations of their assumption eventually will result in unreliable findings. Can we imagine the degree of the damage done to the research due to this mistake? However, most researchers are not aware of the seriousness of the error because they are only the users of the statistical methods. Most quantitative researchers, especially in the field of business, economics, and social sciences, rely heavily on the classical methods to solve their problems. Continuous use of the classical methods without considering the assumptions will most probably generate erroneous results.

The emergence of alternatives such as robust methods could help to reduce the error and improve the statistical testing regardless of the sample sizes. The need and effectiveness of robust methods have been described in many papers and books since decades ago (e.g. [5], [4] and [13]. Departures from normality originate from two problems, i.e. skewness and the existence of outliers. These problems could be remedied by using transformation such as exponential, logarithm and others but sometimes, even after the transformation, problems with non normal data still occur. Simple transformations of the data such as by taking logarithm can reduce skewness but not for complex transformations such as the class of Box-Cox transformations [18]. However, problems due to outliers still exist. According to [18], a simple transformation can alter skewed distributions to make them more symmetrical, but the approach does directly eliminate outliers.

In our study, we would like to suggest a statistical procedure that is known to be able to handle the problems of nonnormality. Known as the modified F_t statistics, this procedure is categorized under robust statistics. Robust statistics combine the virtues of both parametric and nonparametric approaches. In nonparametric inference, few assumptions are made regarding the distribution from which the observations are drawn. In contrast, the approach in robust inference is different wherein there is a working assumption about the form of the distribution, but we are not entirely convinced that the assumption is true. Robustness theories can be viewed as stability theories of statistical inference and signify insensitivity to small deviations from the assumptions [5]. What is

[*] Corresponding author e-mail: zahayu@uum.edu.my

desired is an inference procedure, which in some sense does almost as well as if the assumption is true, but does not perform much worse within a range of alternatives to the assumption. The theories of robustness consider neighborhoods of parametric models and thus belong to parametric statistics. A robust procedure usually adopts what might be called an ”applied parametric viewpoint”, which according to [5] uses a parametric model. This model is hopefully a good approximation to the true underlying situation, but we cannot assume that it is exactly correct. Frequently in discussions of robustness, the assumed distribution (probability density function) is normal; therefore, the type of robustness of interest is ”robustness to non-normality”.

The proposed procedure of modified F_t to be adopted in this study is among the latest procedures in robust statistics, was proposed by [10]. This procedure is for testing the equality of the central tendency measures for J groups with $H_0 : \theta_1 = \theta_2 = \ldots = \theta_J$, where θ_J is the central tendency parameter corresponding to distribution $F_J : J = 1, 2, \ldots, J$. Modified F_t uses trimmed mean as the central tendency measures.

2 Methods

In this section, we discussed on the modified F_t method, which combines F_t statistics with scale estimators suggested by [14].

2.1 F_t Statistics

The original Ft statistic or trimmed F statistic was introduced by [8]. This statistical procedure is able to handle problems with sample locations when non normality occurs but the assumption of homogeneity of variances still applies. This new statistic is easy to compute and is used as an alternative to the classical F method involving one-way independent group design.

To further understand the F_t method, let

$$X_{(1)j}, X_{(2)j}, \ldots, X_{(n_j)j}$$

be an ordered sample of group j with size n_j.

We calculate the trimmed mean of group j by using:

$$\overline{X}_{tj} = \frac{1}{n_j - g_{1j} - g_{2j}} [\sum_{i=g_{ij}+1}^{n_j - g_{2j}} X_{(i)j}]$$

where

= number of observations $X_{(i)j}$ such that g_{ij} that $(X_{(i)j} - \hat{M}_j) < -2.24$(scale estimator),

where

number of observations $X_{(i)j}$ such that g_{2j} that $(X_{(i)j} - \hat{M}_j) > 2.24$(scale estimator),

$\hat{M}_j$ = median of group j, and

scale estimator = MAD_n, T_n or LMS_n.

For the equal amounts of trimming in each tail of the distribution, the Winsorized sum of squared deviations is defined as

$$SSD_{tj} = (g_j + 1)(X_{(g_j+1)j} - \overline{X}_{tj})^2 + (X_{(g_j+2)j} - \overline{X}_{tj})^2 + \ldots \\ + (X_{(n_j-g_j-1)j} - \overline{X}_{tj})^2 + (X_{(n_j-g_j)j} - \overline{X}_{tj})^2$$

When allowing different amounts of trimming in each tail of the distribution, the Winsorized sum of squared deviations is then defined as,

$$SSD_{tj} = (g_{1j} + 1)(X_{(g_{1j}+1)j} - \overline{X}_{tj})^2 + (X_{(g_{1j}+2)j} - \overline{X}_{tj})^2 + \ldots \\ + (X_{(n_j-g_{2j}-1)j} - \overline{X}_{tj})^2 + (g_{2j} + 1)(X_{(n_j-g_{2j}+1)j} - \overline{X}_{tj})^2 \\ - \frac{\{(g_{1j})[X_{(g_{1j}+1)j} - \overline{X}_{tj}] + (g_{2j})(X_{(n_j-g_{2j})j} - \overline{X}_{tj})\}^2}{n_j}$$

Note that we used trimmed means in the SSD_{tj} formula instead of Winsorized means.

Hence the trimmed F is defined as

$$F_{t(j)} = \frac{\frac{\sum_{j=1}^{J} (\overline{X}_{tj} - \overline{X}_j)^2}{(J-1)}}{\sum_{j=1}^{J} \frac{SSD_{tj}}{(H-j)}}$$

where J = number of groups,

$h_j = n_j - g_{1j} g_{2j}$

$H = \sum_{j=1}^{J} h_j$

and $\overline{X}_t = \sum_{j=1}^{J} \frac{h_j \overline{H}_{tj}}{H}$

$F_{t(g)}$ will follow approximately an F distribution with $(J-1, H-J)$ degree of freedom.

2.2 Scale Estimator

Let $X = (x_1, x_2, \ldots, x_n)$ be a random sample from any distribution and let the sample median be denoted by $med_i x_i$.

2.2.1 MAD_n

MAD_n is median absolute deviation about the median. Given by

$$MAD_n = b med|x_i - med x_i|$$

with b as a constant, this scale estimator is very robust with best possible breakdown point and bounded influence function. MAD_n is identified as the single most useful ancillary estimate of scale due to its high breakdown property [5]. This scale estimator is simple and easy to compute.

The constant b is needed to make the estimator consistent for the parameter of interest. For example if the observations are randomly sampled from a normal distribution, by including $b = 1.4826$, the MAD_n will estimate σ, the standard deviation. With constant $b = 1$, MAD_n will estimate 0.75 , and this is known as MAD.

2.2.2 T_n

Suitable for asymmetric distribution, [14] proposed Tn, a scale known for its highest breakdown point like MAD_n. However, this estimator has more plus points compared to MAD_n. It has 52% efficiency, making it more efficient thanMAD_n. It also has a continuous and bounded influence function. Furthermore, the calculation of T_n is much easier than the other scale estimators.

Given as

$$T_n = 1.3800\frac{1}{h}\sum_{k=1}^{h}\{\underset{i\neq j}{med}|x_i - x_j|\}_{(k)}$$

where $h = [\frac{n}{2} + 1]$

T_n has a simple and explicit formula that guarantees uniqueness. This estimator also has 50% breakdown point.

2.2.3 LMS_n

LMS_n is also a scale estimator with 50% breakdown point which is based on the length of the shortest half sample as shown below:

$$LMS_n = c'\{\underset{i}{min}|x_{(i+h-1)} - x_{(i)}|\}$$

given $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ are the ordered data and $h = [\frac{n}{2} + 1]$. The default value of c' is 0.7413 which achieves consistency at Gaussian distributions. LMS_n has influence function which is similar to MAD [13] and its efficiency equals to that of the MAD as well [2].

3 Empirical Investigations

Since this paper deals with robust method where sensitivity to small changes is the main concern, manipulating variables could help in identifying the robustness of each method. Four variables (listed below) were manipulated to create conditions which are known to highlight the strengths and weaknesses of the procedure.

(1)Number of Groups: Investigations were done on four unbalanced completely randomized groups design since previous researches have looked into these designs ([9]; [12]; [19]).

(2)Distributional Shape: In investigating the effects of distributional shape on Type I error and power, two types of distribution representing different level of skewness were being considered. The distributions are the standard normal distribution, and the g-and-h distribution with g = 0.5 and h = 0.5, representing zero and extreme skewness respectively. The skewness for the g-and-h distribution with g = 0.5 and h = 0.5 are undefined.

(3)Variance heterogeneity: Variance heterogeneity is one of the general problems in testing the equality of location measures. Therefore, in looking at the effects of this condition to the test, the variances with ratio 1:1:1:36 were assigned to the groups. Although this ratio may seem extreme, ratios similar to this case, and even larger, have been reported in the literature [7].

Pairings of unequal variances and group sizes: Variances and group sizes were positively and negatively paired for comparison. For positive pairings, the group having the largest group observations was paired with the population having the largest group variance, while the group having the smallest number of observations was paired with the population having the smallest group variance. For negative pairings, the group with the largest number of observations was paired with the smallest group variance and the group with the smallest number of observations was paired with largest group variance. These conditions were chosen since they typically produce conservative results for the positive pairings and liberal results for the negative pairings [11].

The random samples were generated using SAS generator RANNOR [15]. The variates were standardized and transformed to g-and-h variates having mean μ_j and variance σ_j^2 . The design specification for four groups is shown in Table 1.

Table 1 Design specification for four groups

	Group sizes				Population variances			
	1	2	3	4	1	2	3	4
+ve	10	15	20	25	1	1	1	36
-ve	10	15	20	25	36	1	1	1

To test the Type I error, the group means were set as (0, 0, 0, and 0) for the four groups and 5000 data sets were simulated for each design.

4 Simulation results

The robustness of a method is determined by its ability in controlling the Type I error. By adopting Bradley's liberal criterion of robustness [1], a test can be considered robust if its empirical rate of Type I error,α , is within the interval 0.5α and 1.5α . If the nominal level is $\alpha = 0.05$, the empirical Type I error rate should be between 0.025 and 0.075. Correspondingly, a procedure is considered to be non-robust if, for any particular condition, its Type I error rate is not within this interval. We chose this criterion since it was widely used by many robust statistic researchers (e.g. [6]; [12]; [16]; [17]) to judge robustness. Nevertheless, for Guo and Luh (2000), if the empirical Type I error rates do not exceed the 0.075 level, the procedure can be considered robust. The best procedure is the one that can produce Type I error rate closest to the nominal (significance) level.

Table 2 Type I error rates

Dist.	Pair.	Methods				
		Ft with *MADn*	*Ft* with *Tn*	*Ft* with *LMSn*	ANOVA	Kruskall Wallis
Normal	+ve	0.0774	0.0780	0.0498	0.0336	0.0448
	-ve	0.3542	0.3196	0.2868	0.2850	0.1158
	Ave	0.2158	0.1988	0.1683	0.1593	0.0803
g = 0.5, h = 0.5	+ve	0.0370	0.0366	0.1542	0.1492	0.0498
	-ve	0.2814	0.2638	0.3000	0.3554	0.1022
	Ave	0.1592	0.1502	0.2271	0.2523	0.0760

For positive pairing, the F_t with LMS_n, ANOVA and Kruskall Wallis showed robust Type I error rates. The best procedure is F_t with LMS_n which produce the nearest Type I error rate to the nominal level. As can be observed in Table 2, under normal distribution, the average Type I error rates for F_t with MAD_n, F_t with T_n, F_t with LMS_n and ANOVA inflate above the 0.1 level. This is due to the large values of Type I error rates when the pairings are negative.

Under extremely skewed distribution, again, the average results for F_t with MAD_n, F_t with T_n, F_t with LMS_n and ANOVA show inflated average Type I error rates due to the negative pairings. In contrast, the Type I error for F_t with MAD_n, F_t with T_n and Kruskall Wallis improved under positive pairing, indicating robustness. But not in the case of F_t with LMS_n and ANOVA which produced the worst result with Type I error for both pairings are above the 0.1 level.

5 Analysis on Real Data

The performance of the modified F_t method was then demonstrated on real data. Four classes (groups) of Decision Analysis course of the 2nd Semester 2010/2011 taught by 4 different lecturers were chosen at random. The final marks were recorded and tested for the equality between the classes. The sample sizes for Class 1, 2, 3 and 4 were 33, 19, 24 and 20 respectively. The result for the descriptive statistics is given in Table 3.

Table 3 Descriptive statistics for each group

Group	*n*	Mean of the marks	Std. Deviation	Std Error	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
1	33	72.07	15.65	2.72	66.53	77.62	7	94
2	19	70.13	9.13	2.10	65.73	74.53	56	90
3	24	73.38	10.75	2.20	68.84	77.91	60	96
4	20	79.21	6.11	1.37	76.35	82.06	68	93

Table 4 : Results of the test statistics using different methods

Methods	p-value
ANOVA	0.0870
Kruskall Wallis	0.0160
F_t with MAD_n	0.0021
F_t with T_n	0.0020
F_t with LMS_n	0.0407

Table 4 shows the results in the form of p-values for each group tested in this study. For comparison purpose, the data were tested using all the five procedures mentioned in this study namely ANOVA, Kruskal Wallis and the modified F_t with robust scale estimator, MAD_n, T_n and LMS_n. As can be observed in Table 4, when testing using ANOVA, the result fails to reject the null hypothesis that the performance for all groups is equal. On the contrary, when using Kruskall Wallis and modified F_t method, the tests show significant results (reject the null hypothesis).

The result indicates that ANOVA fails to detect the difference which exists between the groups. Both the non parametric (Kruskall Wallis) and robust methods (modified F_t) show better detection. F_t with T_n shows the strongest significance ($p = 0.0020$) as compared to the other methods. As shown in the simulation results in Table 2, modified F_t in general produced robust Type I error rates for extremely skewed distribution.

6 Conclusion

goal of this paper is to find alternative procedures in testing location parameter for skewed distribution. Classical method such as ANOVA is not robust to non normality and heteroscedasticity. When the problems of nonnormality and heteroscedasticity occur simultaneously, the Type I error rate will inflate, causing spurious rejections of the null hypotheses and the power of test can be substantially reduced from theoretical values, resulting in undetected differences. Realizing the need for a good statistic in addressing these problems, we integrate the F_t statistic [8] with the highest breakdown scale estimators [14] and these new methods are known as the modified F_t methods.

This paper has shown some improvement in the statistical solution for detecting differences between location parameters. The findings showed that the modified robust procedures, F_t with MAD_n, F_t with T_n, F_t with LMS_n are comparable with Kruskall Wallis in controlling Type I error rates under most conditions. In the analysis on real data, F_t with T_n (p = 0.0020) and F_t with MAD_n (p = 0.0021) showed stronger significance than Kruskall Wallis (p = 0.0160). Even though F_t with LMS_n (p = 0.0407) showed weaker significance than the aforementioned procedures, its performance was proven to be much better than the parametric ANOVA (p = 0.0870) in both simulation study and real data analysis.

Acknowledgement

The authors would like to acknowledge the work that led to this paper is partially funded by the University Research Grant Scheme of the Universiti Utara Malaysia

References

[1] Bradley, J.V. (1978). Robustness?. *British Journal of Mathematical and Statistical Psychology*, 31, 144 - 152.

[2] Grubel, R. (1988). The length of the shorth. *The Annals of Statistics*, 16, 619 - 628.

[3] Guo, J.- H., & Luh, W. - M. (2000). An invertible transformation two-sample trimmed t-statistic under heterogeneity and nonnormality. *Statistic & Probability letters*, 49, 1 -7.

[4] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust statistics*. New York: Wiley.

[5] Huber, P. J. (1981). *Robust statistics*. New York: Wiley.

[6] Keselman, H.J., Kowalchuk, R.K. Algina, J., Lix, L.M., and Wilcox, R.R. (2000) Testing treatment effects in repeated measures designs: Trimmed means and bootstrapping. *British Journal of Mathematical and Statistical Psychology*, 53: 175-191.

[7] Keselman, H.J., Wilcox, R.R., Algina, J., Fradette, K.,Othman, A.R. (2002). A power comparison of robust test statistics based on adaptive estimators. *Journal of Modern Applied Statistical Methods*.

[8] Lee, H & Fung, K.Y. (1985). Behaviour of trimmed F and sine-wave F statistics in one-way ANOVA. *Sankhya:The Indian Journal of Statistics*, 47 (Series B), 186 - 201.

[9] Lix, L.M and 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.

[10] Md Yusof, Z., Othman, A.R. & Syed Yahaya, S.S. (2007a). *Type I error rates of trimmed F statistic*. In proceeding of the 56th Session of the International Statistical Institute (ISI 2007), 22 - 29 August 2007. Lisbon Portugal. (In CD).

[11] Othman,A.R., Keselman, H.J., Padmanabhan, A.R., Wilcox, R.R and Fradette, K (2003). Comparing measures of the ”typical” score across treatment groups. *British Journal of Mathematical and Statistical Psychology*.

[12] 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*, 215 - 234.

[13] Rousseeuw, P. J., & Leroy, A. M. (1987). *Robust regression and outlier detection*. New York: Wiley.

[14] Rousseeuw, P.J.and Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, 88: 1273-283.

[15] SAS Institute Inc. (1999). SAS/IML *User's Guide version* 8. Cary, NC: SAS Institute Inc.

[16] Syed Yahaya, S.S., Othman, A.R. and Keselman, H.J. (2004). *Testing the equality of location parameters for skewed distributions using S1 with high breakdown robust scale estimators*. In M.Hubert, G. Pison, A. Struyf and S. Van Aelst (Eds.), Theory and Applications of Recent Robust Methods Series: Statistics for Industry and Technology (319-328), Birkhauser, Basel.

[17] Wilcox, R.R., Keselman H.J., Muska, J., Cribbie, R. (2000). Repeated measures ANOVA: Some new results on comparing trimmed means and means. *The British Psychological Society*. 53: 69-82.

[18] Wilcox, R. R., & Keselman, H. J. (2003). Modern robust data analysis methods: Measures of central tendency. *Psychological Methods*, 8, 254-274.

[19] Yuen, K.K. (1974). The two-sample trimmed t for unequal population variances. *Biometrika*, 61: 165- 170.

Zahayu Yusof received the MS degree in Statistics from Universiti Kebangsaan Malaysia in 2001, and the PhD degree in Statistics from Universiti Sains Malaysia in 2010. She is currently a senior lecturer in Universiti Utara Malaysia. Her research interests are in the areas of robust statistics

Suhaidah Abdullah received the MS degree in Statistics from Universiti Kebangsaan Malaysia in 2001, and the PhD degree in Statistics from Universiti Utara Malaysia in 2012. She is currently a senior lecturer in Universiti Utara Malaysia. Her research interests are in the areas of robust statistics.

Sharipah Soaad Syed Yahaya received the MS degree in Mathematics from Indiana State University in 1985 and the PhD degree in Statistics from Universiti Sains Malaysia in 2006. She is currently an associate professor in Universiti Utara Malaysia. Her research interests are in the areas of robust statistics.