International Journal of Interdisciplinary Research in Statistics, Mathematics and Engineering
ISSN: 2992-9261| Impact Factor : 7.62
Volume. 11, Number 2; April-June, 2024;
Published By: Scientific and Academic Development Institute (SADI)
8933 Willis Ave Los Angeles, California
https://sadijournals.org/index.php/ijirsme|editorial@sadijournals.org

TESTING THE EQUALITY OF CENTRAL TENDENCY MEASURES USING VARIOUS TRIMMING STRATEGIES

[1]Zahayu Md Yusof and [2]Abdul Rahman Othman
[1]UUM College of Arts and Sciences, Science Quantitative Building, Universiti Utara Malaysia, 06010 UUM Sintok, Kedah, Malaysia.
[2]Pusat Pengajian Pendidikan Jarak Jauh, Universiti Sains Malaysia, 11800 USM Penang, Pulau Pinang, Malaysia.
DOI: https://doi.org/10.5281/zenodo.12800887

Abstract: Ft statistic test is a non-classical method of comparing two or more groups. This statistical procedure is able to handle problems of sample locations when non-normality occurs but the homogeneity of variances assumption still applies. This method is not robust under the existence of variance heterogeneity. To make this test less sensitive when either one or both of the common assumptions are violated, in this study the test is modified and improved by replacing the test's original central tendency measure that is, the fixed symmetric trimmed mean with a predetermined asymmetric trimmed mean and a modified one-step M-estimator (MOM) trimmed mean. The finding shows that when the data is suspected to be extremely skewed, then, it will be advantageous to adopt MOM procedures for homogeneous variance cases. On the other hand, for heterogeneous variances, a trimmed mean which uses predetermined asymmetric trimmed mean should be considered as an alternative, particularly for testing the equality of four groups.

Key words: Heterogeneity, asymmetric, trimmed mean, MOM, robustness.

INTRODUCTION

In recent years, numerous methods for locating treatment effects or testing the equality of central tendency (location) parameters by simultaneously controlling the Type I error and the power to detect treatment effects are being studied. Progress has been made in terms of finding better methods for controlling the Type I error and the power of the test that detects treatment effects in one-way independent group designs (Babu et al., 1999; Othman et al., 2004; Wilcox and Keselman, 2003). Through a combination of impressive theoretical developments, more flexible statistical methods, and faster computers, serious practical problems that seemed insurmountable only a few years ago can now be addressed. These developments are important to applied researchers because they greatly enhance the ability to discover true differences between groups while maximizing the chance of detecting a genuine positive effect.

Let us look at the example of analysis of variance (ANOVA), and the drawbacks of this method when assumptions are not met. ANOVA is one of the most commonly used statistical methods for locating treatment effects in one-way independent group design. Generally, violating the assumptions associated with standard ANOVA method can seriously hamper the ability to detect true differences.

Non-normality and heteroscedasticity are the two usual assumption violations detected in ANOVA. In particular, when these problems occur at the same time, rates of Type I error is usually inflated, thus resulting in spurious rejections of the null hypotheses. They can also substantially reduce the power of a test, resulting in treatment effects going undetected. Reduction in the power to detect differences between groups occurs because the usual population standard deviation (σ) is very sensitive to outliers and will be greatly influenced by their presence. Consequently, the standard error of the mean (σ^2/n) can become seriously inflated, when the underlying distribution has heavy tails (Wilcox and Keselman, 2002). Therefore, the standard error of the F statistics is larger than it should be and power accordingly will be depressed. In order to achieve a good test, one needs to be able to control Type I error and power of test. In other words, neither should power be lost nor Type I error be inflated.

In their efforts to control the Type I error and power rate, investigators looked into numerous robust methods since these methods generally are insensitive to assumptions about the overall nature of the data (e.g. Babu et al., 1999; Keselman et al., 2004; Kulinskaya et al., 2003; Luh and Guo, 1999; Othman et al., 2004). Any small deviations from the model assumptions should only slightly impair the performance, for example, the level of a test should be close to the nominal value calculated at the model, and larger deviations from the model should not cause catastrophe. Robust measures of central tendency such as trimmed means, medians or Mestimators (Huber, 1981; Staudte and Sheather, 1990; Wilcox, 1997) have been considered as alternatives for the usual least squares estimator, that is, the usual least squares mean, in most research recently (Keselman et al., 2004; Luh and Guo, 1999; Wilcox et al., 1998; Wilcox and Keselman, 2002). These measures of central tendency had been shown to have better control over Type I error and power to detect treatment effects (Lix and Keselman, 1998; Othman et al., 2004; Wilcox, 1997; Yuen, 1974). Yuen (1974) found these benefits in the two-group case of trimmed means and Lix and Keselman (1998) demonstrated similar results in the more than twogroup problem. Other investigators, for example, Babu et al. (1999) used median as the central tendency measure when dealing with skewed distribution and Wilcox and Keselman (2003) introduced a modified one-step Mestimator (MOM) as the central tendency measure when testing for treatment effects.

METHODS

This paper focuses on the Ft method with three difference methods of trimming namely (i) fixed symmetric trimming (ii) predetermined asymmetric trimming and (iii) adaptive trimming using MOM. The three trimming methods are different, in the sense that, for the fixed symmetric trimmed mean, data is symmetric and trimming was done equally on both sides of the distribution or no trimming at all. As for predetermined asymmetric trimmed mean, data is either right skewed or left skewed, thus the trimming is done on the skewed side of the distribution. Both of the trimmed mean in (i) and (ii) adopts 15% total amount of trimming. The third, which is the MOM trimmed means need no fixed amount of trimming, but were empirically determined.

Ft statistic

Lee and Fung (1985) introduced a statistic that was able to handle problems with sample locations when the variance for the population is equal. This statistic was named trimmed F statistic, Ft. Their work focused on the best trimming percentages that is able to control Type I error and to provide good power rates. They recommended the trimmed F statistic with 15% symmetric trimming as an alternative to the usual F test especially when the distribution is long tailed symmetric. This method had also been proven to be easy to program.

To further understand the Ft method, let $X_{(1)j}, X_{(2)j}, \ldots, X_{(n_j)j}$ be an ordered sample of group j with size nj and let

$k_j = [g n_j] + 1$

where $[x]$ is the largest integer $\leq x$.

We calculated the g-trimmed mean of group j by using:

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

Where, g_{1j} = number of observations $X_{(i)j}$ such that $(X_{(i)j} - M_j) < -2.24$ (scale estimator), g_{2j} = number of observations $X_{(i)j}$ such that $(X_{(i)j} - M_j) > 2.24$ (scale estimator), M_j = median of group j and the scale estimator is MADn. nj = group sample sizes

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

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

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

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

Note that we used trimmed means in the SSDtj formula instead of Winsorized means.

Hence the g-trimmed F is defined as:

$$F_t(j) = \frac{\sum_{j=1}^{J} (\bar{X}_{tj} - \bar{X}_t)^2 / (J-1)}{\sum_{j=1}^{J} SSD_{tj} / (H - J)}$$

$$F_t(j) = \frac{\sum_{j=1}^{J} (\bar{X}_{tj} - \bar{X}_t)^2 / (J-1)}{\sum_{j=1}^{J} SSD_{tj} / (H - J)}$$

$h^j = n^j - g^{1j} - g^{2j}$,

Where, J = number of groups,

$H = J\ hj \quad X \quad t = J \quad hj\ X\ tj\ /\ H\ j{=}1$ and $j{=}1$

$F^t(g)$ will follow approximately an F distribution with (J – 1, H – J) degree of freedom.

Fixed symmetric trimmed mean

$X(1)\ j \le X(2)\ j \le \ldots \le X(nj\ )\ j$

Let represent the ordered observations associated with the jth group. In order to calculate the 100g% sample trimmed mean, define

XLj = (1 – r)X(k + 1)j + rX(k)j and

XUj = (1 – r) $X(nj-k)\ j$ + r $X(nj-k+1)\ j$

Where, g represents the proportion of observations that are to be trimmed in each tail of the distribution.

k =[gn$_j$]+1

where $[gn^j]$ is the largest integer $^{\le gn\ j}$ and $^{r\ =\ k\ -gn\ j}$.

The jth group trimmed mean is given by

$$\bar{X}_{tj} = \frac{1}{(1-2g)n_j}\left[\sum_{i=k+1}^{n_j-k} X_{(i)j} + r\left(X_{(k)j} + X_{(n_j-k+1)j}\right)\right]$$

Its corresponding sample Winsorized mean is given by

$$\bar{X}_{wj} = \frac{1}{n_j}\left[\sum_{i=k+1}^{n_j-k} X_{(i)j} + k(X_{Lj} + X_{Uj})\right]$$

The g-Winsorized sum of squared deviations is then calculated as

$$SSD = \sum_{i=k_j+1}^{n_j-k_j} (X_{ij} - \bar{X}_{wgj})^2 + k\left[(X_{k_j+1,j} - \bar{X}_{wgj})^2 + (X_{n_j-k_j,j} - \bar{X}_{wgj})^2\right]$$

Predetermined asymmetric trimmed mean

Let $X^{(1)j}, X^{(2)j}, \ldots, X^{(nj)j}$ be an ordered sample of group j with size nj. According to Reed (1998), the γ-trimmed mean (X^{tj}) is defined as

$$\bar{X}_{tj} = \frac{1}{n_j(1-2\gamma)}\left[\sum_{i=k_j+1}^{n-k_j} X_{(i)j} + (k_j - \gamma n_j)\left(X_{(k)j} + X_{n_j-k_j-1}\right)\right]$$

where γ is a proportion that has been trimmed from each tail.

Reed and Stark (1996) proposed an adaptive linear estimator that has the capability of asymmetric trimming. They defined their approach as follows:

Set the value for the total amount of trimming from the sample, γ . Determine the proportion to be trimmed from the lower end of the sample (γ1) by the proportion

$$\gamma_l = \gamma \frac{UW_x}{UW_x + LW_x},$$

Where, UW_{Q_1} =U0.2 − L0.2 and LW_{Q_1} =U0.5 − L0.5

The upper trimming proportion is then given by $\gamma^u = \gamma - \gamma^1$

The Winsorized variance $S^{tj}2$ is defined as:

$\frac{1}{(n_j-1)(1-2\gamma)^2}$ $Stj2 \quad \bar{} = nj - k \quad \bar{} \quad j\ (X(1)\,j \quad \bar{} \; -X\,tj\,)2 + k\,j\ (X(k)\,j - X\,tj\,)2 + k\,j\ (X\,nj - k\,j + 1 - X\,tj\,)2$
$i = k_j + 1$

$k^j = [\gamma n^j] + 1.$

where Yi is the ith ordered observation and observations $X(i)\,j$ such that$(X(i)\,j - M\,j\,) > 2.24$ MADn,

M^j = median of group j and nj = group sample sizes

MOM trimmed mean

$X^{(1)j}, X^{(2)j}, ..., X^{(nj)j}$

Let be an ordered sample of group j with size nj. We calculated the MOM trimmed mean of group j by using:

$1 \quad nj - g2_j$

$X(i)\,j \quad \bar{X}_{tj} = \overline{\qquad\qquad}$

$n\,j - g1j - g2\,j \quad i = g1\,j + 1$

g^{1j} =number of observations $X^{(i)j}$ such

Where,

$(X^{(i)j} - M^j) < -2.24$ MADn, g^{2j} =number of

that

Table 1. Some properties of the *g*- and *h*- distribution.

G	***h***	**Skewness**	**Kurtosis**	**Shape**
0.0	0.0	0.0	3.0	Normal
0.5	0	1.81	9.7	Skewed normal-tailed
0.5	0.5	120.1	18393.6	Skewed heavy-tailed

The value 2.24 was suggested by Wilcox and Keselman (2003). This value is used when checking the extreme values, as it has a reasonably small standard error when sampling from normal distribution. MADn is a value of MAD/0.6745. For the equal amounts of trimming in each tail of the distribution, the Winsorized sum of squared deviations is defined as:

$SSDtj = (gj+1)(X(gj+1)\,j - \bar{X}\,tj\,)2 + (X(gj+2)\,j - X\,tj\,)2 + ...$
$+(X(nj - gj - 1)\,j - X\,tj\,)2 + (gj+1)(X(nj - gj\,)\,j - X\,tj\,)2$

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

$SSDtj = (g1j+1)(X\,(g1\,j+1)j - X\,tj\,)2 + (X\,(g1\,j+2)j - X\,tj\,)2 + ...$
$+(X\,(nj - g2\,j - 1)j - X\,tj\,)2 + (g2\,j+1)(X\,(n\,j - g2\,j\,)j - X\,tj\,)2$
$-\{(g1j\,)[X\,(g1\,j+1)j - X\,tj\,] + (g2\,j\,)[X\,(n\,j - g2\,j\,)j - X\,tj\,]\}2\,/\,n\,j$

Note that we used trimmed means in the SSDtj formula instead of Winsorized means

Empirical investigation

In studying the robustness of the procedures, four variables were manipulated, creating conditions which are known to highlight the strengths and weaknesses of the tests. The four variables were: (1) balanced and unbalanced sample sizes, (2) variance heterogeneity, (3) pairing of group variances and group sample sizes, and (4) types of distributions.

To examine the effect of sample sizes on Type I error rates of the investigated procedures, balanced and unbalanced sample sizes were assigned to the case of four groups (J = 4). Total sample sizes for J = 4 was set at 60 and 80. For the unbalanced sample sizes, each of the four groups were arbitrarily assigned different numbers of observations (nj), namely, n1 = 12, n2 = 14, n3 = 16 and n4 = 18 for N = 60. When N was set at 80, the sample sizes were distributed as n1 = 10, n2 = 20, n3 = 20 and n4 = 30. For the convenience of comparison, the total for the balanced sample sizes were kept constant at 60 and 80. Such that for N = 60, the number of observations for each group was pegged at 15 (that is, n1 = 15, n2 = 15, n3 = 15 and n4 = 15), and for N = 80, it was pegged at 20 (that is, n1 = 20, n2 = 20, n3 = 20 and n4 = 20). To investigate the effect of variance heterogeneity on Type I error rates, variances with a 1:1 and 1:36 ratio were assigned to the groups. To evaluate the robustness of the procedures in relation to the nature of the pairings, each of the proposed procedures was examined under two types of pairings, namely, positive and negative. In investigating the effects of distributional shape on Type I error, three types of distributions representing different levels of skewness (that is normal, skewed normal-tailed and skewed heavy-tailed) using the g- and h- distributions were considered. The g- and h- distributions were modified from normal distribution with constant g controlling the value of skewness and h controlling the value of kurtosis. The level of skewness and kurtosis will increase as the value of g and h increase, respectively. The data is symmetric when g = 0 and h = 0. The values of (g,h) used in this study are (0,0), (0.5,0) and (0.5,0.5). Table 1 summarizes the skewness and kurtosis values for the three selected situations (Wilcox, 1997).

This study was based on simulated data. For data generation, SAS function RANNOR (SAS Institute, 1999) was used to obtain pseudo-random standard normal variates. Observations of the g- and h- distribution were generated by converting the standard normal variates using the following equation:

$$Y_{ij} = \frac{\exp(gZ_{ij})-1}{g}\exp(hZ_{ij}^2/2) \quad \text{for } g \; 0 \qquad \text{and}$$

$$Y_{ij} = Z_{ij}\exp(hZ_{ij}^2/2) \text{ for } g = 0.$$

In examining the Type I error rates, the group location measures were set to zero. For each condition examined, 5000 data sets were generated. The nominal level of significance was set at α = 0.05.

RESULTS

In order to evaluate the particular conditions under which a test is robust, the Bradley's liberal criterion of robustness (Bradley, 1978) was employed. According to this criterion, in order for a test to be considered robust, its empirical rate of Type I error must be within the interval of 0.025 and 0.075. Type I error rates are considered liberal when they are above the 0.075 limit while those below the 0.025 limit are considered conservative.

Tables 2, 3, 4 and 5 show the empirical results of Type I error rates for Ft statistic using fixed 15% symmetric trimmed mean, , predetermined asymmetric trimmed mean and MOM trimmed mean for different sample sizes and types of variances. The bold entries denote the robust result. When the sample sizes are equal, and variance are homogenous, the overall performance of the procedures, which is represented by the "Average" Type I error rates shows that, with the exception of the MOM procedure, all the other procedures are robust as their values fall within Bradley's robust criterion interval for N =

60. The (0.0424) established itself as the best procedure with its Type I error rates being nearest to the nominal level, followed closely by the asymmetric (0.0288) procedures. As the total sample size increases, the number of robust values decreases. With regard to distributional shape, the ν procedure is the best ranked for all types of distribution regardless of N. The asymmetric procedures also performed exceptionally well under normal and skewed normal-tailed distribution, but failed to perform under skewed heavy-tailed. In contrast, the MOM procedure does not perform under normal and skewed normal-tailed distribution but displays a remarkable performance when tested under skewed heavy-tailed distribution. As the N increases, the performance gets better for MOM procedure. MOM procedure gives better performance than the ν procedure when the distribution is skewed heavy-tailed.

When the condition of variance changes to heterogeneous while retaining the equal sample sizes, the procedures using predetermined asymmetric trimmed mean resulted in Type I error rates within the robustness criterion for both N = 60 and N = 80 under normal and skewed heavy-tailed distribution (Table 3). As for skewed normal-tailed distributions, none of the procedures satisfies the Bradley's liberal criterion of robustness.

Looking across the Table 4 values, every entry in the first column that belongs to the ν, is highlighted in bold, which reflects that the procedure has good control of Type I error rates across the three types of distributions for both sample sizes. The results improved as the total sample size increases. The results also show that asymmetric procedure is robust when the distributions are normal and skewed normal-tailed. The MOM is robust when the distribution is skewed heavy-tailed regardless of N. The results of the investigation on unequal sample sizes and heterogeneous variances are presented in Table 5. For this case, there is an additional column for the pairing category. Positive pairing refers to the case in which the largest sample size is associated with the largest group variance and the smallest sample size is associated with the smallest group variance. While negative pairing refers to the case in which the smallest sample size is associated with the largest group variance, and the largest sample size is associated with the smallest group variance.

As shown in Table 5, the Ft statistic with νare robust across the three types of distributions under positive pairing only for both total sample sizes. Like the Ft with ν, the Ft with asymmetric trimming also robust when the pairing is positive throughout the three distributions for N = 60, as the number of total sample size increases to N = 80, the number of robust Type I error rates dwindles to only one, that is when the pairing is positive under normal distribution. While for Ft with MOM, none of the Type I error is robust when N = 60 but as N increases to 80, the procedure becomes robust under normal and skewed normal-tailed distribution but this happened for positive pairing only. None of the procedures seem to be robust when the pairing is negative. The overall result based on the "Grand Average" shows that, the Type I error rates are within the Bradley's interval only for Ft with asymmetric estimator.

Table 2. Type I error rates (Equal sample sizes with homogenous variances).

	***N* = 60 (15, 15, 15, 15)** Variances (1:1:1:1)			***N* = 80 (20, 20, 20, 20)** Variances (1:1:1:1)		
	ν (15%)	**Asymmetric (15%)**	***MOM***	**ν (15%)**	**Asymmetric (15%)**	***MOM***
Normal	0.0476	0.0314	0.1260	0.0488	0.0196	0.1204
Skewed normal-tailed	0.0446	0.0404	0.1686	0.0478	0.0140	0.1592

	$\hat{v}$ (15%)	Asymmetric (15%)	*MOM*	$\hat{v}$ (15%)	Asymmetric (15%)	*MOM*
Normal	0.1084	0.1028	0.1740	0.1096	0.0746	0.1664
Skewed normal- tailed	0.1078	0.0994	0.2106	0.1118	0.0760	0.2048
Skewed heavy-tailed	0.0916	0.0508	0.1184	0.0968	0.0196	0.1026
Average	0.1026	0.0843	0.1676	0.1061	0.0567	0.1579

Skewed heavy-tailed	0.0350	0.0146	0.0588	0.0384	0.0006	0.0472
Average	0.0424	0.0288	0.1178	0.0450	0.0114	0.1089

Table 3. Type I erroral sample sizesheterogeneous rates (Equ with varia nces).

N = 60 (15, 15, 15, 15)	*N* = 80 (20, 20, 20, 20)
Variances (1:1:1:36)	Variances (1:1:1:36)

Table 4. Type I error rates (Unequal sample sizes with homogeneous variances).

	$\hat{v}$ (15%)	Asymmetric (15%)	*MOM*	$\hat{v}$ (15%)	Asymmetric (15%)	*MOM*
Normal	0.0476	0.0312	0.1194	0.0532	0.0224	0.1162
Skewed normal- tailed	0.0458	0.0382	0.1632	0.0530	0.0272	0.1508
Skewed heavy-tailed	0.0342	0.0164	0.0572	0.0474	0.0170	0.0542
Average	0.0425	0.0286	0.1133	0.0512	0.0222	0.1071

N = 60 (12, 14, 16, 18)	*N* = 80 (10, 20, 20, 30)
Variances (1:1:1:1)	Variances (1:1:1:1)

Table 5. Type I error rates (Unequal sample sizes with heterogeneous variances).

		N = 60 (12, 14, 16, 18) Variances (1:1:1:36)			*N* = 80 (10, 20, 20, 30) Variances (1:1:1:36)		
		(15%)	**Asymmetric (15%)**	***MOM***	**(15%)**	**Asymmetric (15%)**	***MOM***
Normal	Positive	0.0696	0.0574 0.1676	0.1208	0.0356	0.0284 0.2622	0.0652
	Negative	0.1434		0.2440	0.2692		0.3812
Average		0.1065	0.1125	0.1824	0.1524	0.1453	0.2232
Skewed normal- tailed	Positive	0.0688	0.0626	0.1482	0.0358	0.0202	0.0878
	Negative	0.1408	0.1510	0.2796	0.2720	0.2696	0.4184
Average		0.1048	0.1068	0.2139	0.1539	0.1449	0.2531
Skewed heavy-tailed	Positive	0.0526	0.0332	0.0702	0.0292	0.0056	0.0284
	Negative	0.1172	0.0884	0.1720	0.2594	0.1508	0.3112
Average		0.0849	0.0608	0.1211	0.1443	0.0782	0.1698
Grand Average		0.0987	0.0934	0.1725	0.1502	0.1228	0.2154

DISCUSSION AND CONCLUSION

Our goal is to search for some alternative methods in testing the equality of central tendency (location) measures for skewed distributions. For the homogeneous variances, regardless of sample design (balanced or unbalanced), the Ft procedures with fixed symmetric trimmed mean, ν performed so well under normal-tailed distribution, but lost control over Type I error under extreme conditions. When researchers suspect that their data is extremely skewed, in a manner similar to the characteristics of the g- and h- distribution (g = 0.5 and h = 0.5), then clearly, it will be advantageous to adopt MOM procedures. As for heterogeneous variances, a predetermined asymmetric trimmed means which used 15% total amount of trimming should be considered as an alternative particularly for testing the equality of four groups.

These modified methods may serve as alternatives to some other robust statistical methods which are unable to handle either the problem of non-normality, variance heterogeneity or unbalanced design. This study may generate ideas for future research in robust methods simultaneously contributes to filling the gaps in the literature in this field.

ACKNOWLEDGEMENT

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

REFERENCES

Babu JG, Padmanabhan AR, Puri MP (1999). Robust one-way ANOVA under possibly non-regular conditions. Biometrical J., 41(3), 321 – 339.

Bradley JV (1978). Robustness? British J. Math. Stat. Psychol., 31: 144 - 152.

Huber PJ (1981). Robust statistics. New York: Wiley.

Keselman HJ, Othman AR, Wilcox RR, Fradette K (2004). The new and improved two-sample t test. Psychol. Sci., 15: 47-51.

Kulinskaya E, Staudte RG, Gao H (2003). Power approximations in testing for unequal means in a one-way ANOVA weighted for unequal variances. Commun. Stat. – Theory Methods, 32: 2353 – 2371.

Lee H, Fung KY (1985). Behaviour of trimmed F and sine-wave F statistics in one-way ANOVA. Sankhya: The Indian J. Stat., 47 (Series B): 186 - 201.

Lix LM, Keselman HJ (1998). To trim or not to trim: Tests of location equality under heteroscedasticity and nonnormality. Educ. Psychol. Measure., 58(3): 409 – 429.

Luh WM, Guo JH (1999). A powerful transformation trimmed mean method for one-way fixed effects ANOVA models under non- normality and inequality of variances. British J. Math. Stat. Psychol., 52: 303-320.

Othman AR, Keselman HJ, Padmanabhan AR, Wilcox RR, Fredette K (2004). Comparing measures of the 'typical' score across treatment groups. British J. Math. Stat. Psychol., 57: 215-234.

Reed III JF, Stark DB (1996). Hinge estimators of location: Robust to asymmetry. Comput. Methods Programs Biomed., 49: 11 – 17.

Reed III JF (1998). Contributions to adaptive estimation.

25: 651 – 669.

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

Staudte RG, Sheather, SJ (1990). Robust estimation and testing. New York: Wiley.

Wilcox RR (1997). Introduction to robust estimation and hypothesis testing. San Diego, CA: Academic Press.

Wilcox RR, Keselman HJ, Kowalchuk, RK (1998). Can tests for treatment group equality be improved? The bootstrap and trimmed means conjecture. British J. Math. Stat. Psychol., 51: 123 – 134.

Wilcox RR, Keselman HJ (2002). Power analysis when comparing trimmed means. J. Modern Appl. Stat. Methods, 1(1): 24-31.

Wilcox RR, Keselman HJ (2003). Modern robust data analysis methods: Measures of central tendency. Psychol. Methods, 8: 254-274.

Yuen KK (1974). The two-sample trimmed t for unequal population variances. Biometrika, 61: 165 – 170.