

Type I Error Rates of F_t Statistic with Different Trimming Strategies for Two Groups Case

Zahayu Md Yusof, Suhaida Abdullah & Sharipah Soaad Syed Yahaya

UUM College of Arts and Sciences, Science Quantitative Building, Universiti Utara Malaysia
06010 UUM Sintok, Kedah, Malaysia

E-mail: zahayu@uum.edu.my, suhaida@uum.edu.my, sharipah@uum.edu.my

Abdul Rahman Othman

Pusat Pengajian Pendidikan Jarak Jauh, Universiti Sains Malaysia
11800 USM Penang, Pulau Pinang, Malaysia

E-mail: oarahman@usm.my

Received: May 19, 2011

Accepted: June 30, 2011

doi:10.5539/mas.v5n4p236

Abstract

When the assumptions of normality and homoscedasticity are met, researchers should have no doubt in using classical test such as t -test and *ANOVA* to test for the equality of central tendency measures for two and more than two groups respectively. However, in real life we do not often encounter with this ideal situation. A robust method known as F_t statistic has been identified as an alternative to the above methods in handling the problem of nonnormality. Motivated by the good performance of the method, in this study we proposed to use F_t statistic with three different trimming strategies, namely, i) fixed symmetric trimming (10%, 15% and 20%), ii) fixed asymmetric trimming (10%, 15% and 20%) and iii) empirically determined trimming, to simultaneously handle the problem of nonnormality and heteroscedasticity. To test for the robustness of the procedures towards the violation of the assumptions, several variables were manipulated. The variables are types of distributions and heterogeneity of variances. Type I error for each procedures were then be calculated. This study will be based on simulated data with each procedure been simulated 5000 times. Based on the Type I error rates, we were able to identify which procedures (F_t with different trimming strategies) are robust and have good control of Type I error. The best procedure that should be taken into consideration is the F_t with $MOM - T_n$ for normal distribution, 15% fixed trimming for skewed normal-tailed distribution and $MOM - MAD_n$ for skewed leptokurtic distribution. This is because, all of the procedures produced the nearest Type I error rates to the nominal level.

Keywords: Type I error, Trimming, Robust scale estimator, Hinge estimator, Heterogeneity

1. Introduction

The classical tests of group equality such as the t test and analysis of variance (*ANOVA*) are always misrepresented due to variance heterogeneity and nonnormality. When the problems of nonnormality and variance heterogeneity arise simultaneously, Type I error rates are usually inflated resulting in spurious rejection of null hypotheses and reduction in the power of the test statistics. These problems could be overcome by (i) applying transformation to the data and proceed with the procedures or (ii) selecting an alternative procedure which is insensitive (robust) to the problems. When data are non normal and variances are heterogeneous, it is often possible to transform the data so that the new scores are normal with equal variances. Although transformation is capable of normalizing skewed data, the method still has some drawbacks. According to Wilcox (2002), transformation (e.g. square root of the mean and log of the mean) overlooks the inferences of real data. It is also unable to eliminate the influence of outliers and sometimes even a complicated transformation fail to normalize the data. For that reason, this study will look into the second approach in alleviating the problem of non normality and heterogeneity of variances simultaneously. Lix and Keselman (1998) stated that tests which are sensitive to the combined effects of nonnormality and variance heterogeneity can be obtained by substituting robust measures of location and scale such as trimmed means and Winsorized variances in place of the usual means and variances respectively. Trimmed mean is a good measure of location because its standard error is less affected by departures from normality. This is due to the fact that the extreme values or outliers are removed (Lix & Keselman, 1998). It is a popular robust estimator used by many researchers because of its ability in controlling Type I error rate (Luh, 1999; Keselman, Wilcox, Taylor & Kowalchuk, 2000; Locker, 2001; Luh & Guo, 2005). A few examples of robust methods which uses trimmed mean are the Welch test (Welch, 1951), the James test (James, 1951) and the Alexander-Govern test (Alexander & Govern, 1994). Trimmed mean summarizes data when trimming is carried out. By using the trimmed means, the effect of the tails of the distribution is reduced by their removal based on the trimming percentage that has to be stated in advanced (fixed amount). Even though trimmed mean has a good control of Type I error rate, the trimming is done regardless of the types of distribution due to the percentage of trimming that

is set prior to the facts whether the outliers are presence or not. It will be a gross mistake to eliminate data which are not outliers such as in a normally distributed data. There is extensive literature regarding this trimming method that uses the fixed amount of symmetric trimming. Among them are Lee and Fung (1985), Keselman, Wilcox, Othman and Fradette (2002), and Wilcox (2003).

To avoid unnecessary trimming, an alternative to a fixed trimmed mean is the modified one-step M estimator (hereafter MOM). This estimator empirically trimmed extreme data based on the types (shape) of distribution (Othman, Keselman, Padmanabhan, Wilcox & Fradette, 2004). If we have skewed distributions then the amount of trimming on both tails should not be the same. More should be trimmed from the skewed tail unlike trimmed mean which trimmed the data regardless of the shape of the tails. Another approach of trimming which was investigated by Keselman, Wilcox, Lix, Algina and Fradette (2007) used fixed asymmetric trimming. This trimming approach uses hinge estimators proposed by Reed and Stark (1996) to determine the suitable amount of trimming on each tail of a distribution. However, their method still adopts the fixed trimming percentages.

Motivated by the good performance of trimmed mean, this study will investigate on the three approaches of trimmed mean mentioned above i.e. (i) fixed (percentage) symmetric trimming, (ii) automatic trimming and (iii) fixed asymmetric trimming. These trimmed means will be used as the central tendency measures for a robust test statistic known as F_t statistic. The performance of these methods in terms of type I error rates for the case of two groups are determined and compared.

2. Method

Three methods of trimming namely the fixed symmetric trimming, automatic trimming and fixed asymmetric trimming were examined in this study. The fixed symmetric and asymmetric trimming adopted 10%, 15% and 20% total amount of trimming while the automatic trimming trimmed data based on different scale estimators i.e. MAD_n , T_n , and LMS_n ,

Each trimmed mean generated through different criteria of trimming was used as the central tendency measure for F_t statistic. Altogether, this study produced nine different F_t procedures. Each of these procedures was then compared under conditions of normality and non-normality represented by the g - and h - distributions.

2.1 F_t statistic

Lee and Fung (1985) introduced a statistical procedure that is able to handle problems with sample locations when nonnormality occurs but the homogeneity of variances assumption still applies. This statistic was named trimmed F statistic, F_t . Their work focused on the best trimming percentages used to produce trimmed means which are able to control Type I error and to provide good power rates of the statistical procedure.

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 is simple and easy to program.

To further understand the F_t method, let

$X_{(1)j}, X_{(2)j}, \dots, X_{(n_j)j}$ be an ordered sample of group j with size n_j and let $k_j = \lfloor gn_j \rfloor + 1$ where $\lfloor x \rfloor$ is the largest integer $\leq x$.

Hence the g -trimmed F is defined as

$$F_t(g) = \frac{\sum_{j=1}^J (\bar{X}_{gj} - \bar{X}_t)^2 / (J-1)}{\sum_{j=1}^J SSD_{gj} / (H-J)}, \quad \text{where} \quad [1]$$

$$J = \text{number of groups}, \quad h_j = n_j - g_{1j} - g_{2j}, \quad H = \sum_{j=1}^J h_j \quad \text{and} \quad \bar{X}_t = \sum_{j=1}^J h_j \bar{X}_{gj} / H.$$

g_j represents the proportion of observations in j^{th} group that are to be trimmed in each tail of the distribution.

\bar{X}_{gj} = the j^{th} group trimmed mean, and SSD_{gj} = the g -Winsorized sum of squared deviations. $F_t(g)$ will follow approximately an F distribution with $(J-1, H-J)$ degrees of freedom.

2.2 Fixed symmetric trimmed mean

Let $X_{(1)j} \leq X_{(2)j} \leq \dots \leq X_{(n_j)j}$ represent the ordered observations associated with the j^{th} group. In order to calculate the 100 $g\%$ sample trimmed mean, we define

$$X_{Lj} = (1-r)X_{(k+1)j} + rX_{(k)j} \quad \text{and} \quad X_{Uj} = (1-r)X_{(n_j-k)j} + rX_{(n_j-k+1)j}$$

where

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

$$k = \lfloor gn_j \rfloor + 1 \quad \text{where} \quad \lfloor gn_j \rfloor \text{ is the largest integer } \leq gn_j \quad \text{and} \quad r = k - gn_j.$$

The j^{th} group trimmed mean is given by

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

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

$$SSD_g = \sum_{i=k+1}^{n-k} (X_{ij} - \bar{X}_{gg})^2 + k_j \left[(X_{k+1,j} - \bar{X}_{gg})^2 + (X_{n-k,j} - \bar{X}_{gg})^2 \right] \tag{3}$$

2.3 Automatic trimmed mean

Let $X_{(1)j}, X_{(2)j}, \dots, X_{(n)j}$ be an ordered sample of group j with size n_j . MOM trimmed mean of group j is calculated by using:

$$\bar{\theta}_g = \frac{1}{n_j - g_{1j} - g_{2j}} \left[\sum_{i=g_{1j}+1}^{n-g_{2j}} X_{(i)j} \right] \tag{4}$$

where

g_{1j} = number of observations $X_{(i)j}$ such that $(X_{(i)j} - \hat{M}_j) < -K$ (scale estimator),

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

\hat{M}_j = median of group j and the scale estimator can be either MAD_n, T_n or LMS_n .

$K = 2.24$ (multiplier of scale estimator)

n_j = group sample sizes

The value $K = 2.24$ was suggested by Wilcox and Keselman (2003) in place of the multiplier of the scale estimator in the above criteria. They adjusted the K value so that efficiency is good under normality especially for small sample sizes. They found that, by using simulation with 10,000 replications, the efficiency of $\bar{\theta}_g$ (the standard error of the sample mean divided by the standard error of $\bar{\theta}_g$) is approximately 0.9 for $n_1 = n_2 = n_3 = n_4 = n_5 = 20$ with $K = 2.24$. $\bar{\theta}_g$ was arrived at using MAD_n . We conducted a similar simulation study on $\bar{\theta}_g$ using robust scale estimators T_n and LMS_n , and found that the efficiencies are approximately 0.83 and 0.91, respectively. Hence, we kept the value of 2.24 in our selection criteria. Note that 2.24 is approximately equal to the square root of the 0.976 quantile of a chi-square distribution with one degree of freedom (Wilcox & Keselman, 2003). Indicating that, it is also suitable for skewed distribution.

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

$$SSD_g = (g_j + 1)(X_{(g_j+1)j} - \bar{X}_g)^2 + (X_{(g_j+2)j} - \bar{X}_g)^2 + \dots + (X_{(n-g_j-1)j} - \bar{X}_g)^2 + (g_j + 1)(X_{(n-g_j)j} - \bar{X}_g)^2 \tag{5}$$

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

$$SSD_g = (g_{1j} + 1)(X_{(g_{1j}+1)j} - \bar{\theta}_g)^2 + (X_{(g_{1j}+2)j} - \bar{\theta}_g)^2 + \dots + (X_{(n-g_{1j}-1)j} - \bar{\theta}_g)^2 + (g_{2j} + 1)(X_{(n-g_{2j})j} - \bar{\theta}_g)^2 - \left[(g_{1j})(X_{(g_{1j}+1)j} - \bar{\theta}_g) + (g_{2j})(X_{(n-g_{2j})j} - \bar{\theta}_g) \right]^2 / n_j \tag{6}$$

2.3.1 MAD_n

MAD_n is the median absolute deviation about the median. It demonstrates the best possible breakdown value of 50%, twice as much as the interquartile range and its influence function is bounded with the sharpest possible bound among all scale estimators (Rousseeuw & Croux, 1993).

This robust scale estimator is given by

$$MAD_n = b \text{ med}_i |x_i - \text{med}_j x_j| \tag{7}$$

where the constant $b = 1.4826$ is needed to make the estimator consistent for the parameter of interest, $x_i = x_1, x_2, \dots, x_n$ and $i > j$

However, there are drawbacks in this scale estimator. The efficiency of MAD_n is very low with only 37% at Gaussian distribution. Rousseeuw and Croux (1993) carried out a simulation on 10,000 batches of Gaussian observations to verify the efficiency gain at finite samples. They compared the variance of the standard deviation with the variance of MAD_n based on the finite samples. MAD_n also takes a symmetric view on dispersion and does not seem to be a natural approach for problems with asymmetric distributions.

2.3.2 T_n

Suitable for asymmetric distribution, Rousseeuw and Croux (1993) proposed T_n , 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 than MAD_n . It also has a continuous and bounded influence function. Furthermore, the calculation of T_n is easy.

Given as

$$T_n = 1.3800 \frac{1}{h} \sum_{k=1}^h \left\{ med_{j \neq i} |x_i - x_j| \right\}_{(k)} \quad \text{where } h = \left[\frac{n}{2} \right] + 1 \quad [8]$$

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

2.3.3 LMS_n

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

$$LMS_n = c' \min_i |x_{(i+h-1)} - x_{(i)}| \quad [9]$$

given $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ are the ordered data and $h = \left[\frac{n}{2} \right] + 1$. The default value of c' is 0.7413 which achieves

consistency at Gaussian distributions. LMS_n has an influence function the same as MAD (Rousseeuw & Leroy, 1987) and its efficiency equals that of the MAD as well (Grubel, 1988).

2.4 Adaptive trimmed mean

This adaptive trimmed mean uses hinge estimator HQ_1 (Reed & Stark, 1996) in order to adjust the trimming process that suits the shape of data distribution. Keselman *et al.* (2007) successively improved Welch test using this adaptive trimmed mean in controlling Type I error rates.

The adaptive trimmed mean is calculated as

$$m(\gamma_l, \gamma_u) = \frac{1}{h} \sum_{i=g_1+1}^{n_j-g_2} Y_i \quad [10]$$

where $g_l = [n_j \gamma_l]$, $g_u = [n_j \gamma_u]$, $h = n_j - g_l - g_u$, γ_l = lower trimming percentage, γ_u = upper trimming percentage and n_j is the sample size. The percentage of lower and upper trimming identified using hinge estimator HQ_1 (Reed & Stark, 1996). However the total percentage of trimming is predetermined just like the usual trimmed mean.

To define the lower and upper trimming percentage, let consider an ordered sample J , L_α is the mean of the smallest $[\alpha n]$ observations, where $[\alpha n]$ denotes αn rounded down to the nearest integer, while U_α is the mean of the largest $[\alpha n]$ observations. As for example, let $\alpha = 0.05$, therefore $L_{0.05}$ is the mean of the smallest 0.05n observations.

The measurement of Q_1 is defined as

$$Q_1 = \frac{U_{0.2} - L_{0.2}}{U_{0.5} - L_{0.5}} \quad [11]$$

Q_1 classifies whether a symmetric distribution has light (for $Q_1 < 2$), medium (for $2.6 < Q_1 \leq 3.2$) or heavy (for $Q_1 > 3.2$) tail. It is a location free statistic and uncorrelated with other location statistics. Reed and Stark (1996) defined a general scheme of their approach based on the former definitions of tail length as follows:

- i. Set the total amount of trimming, γ , from the sample.
- ii. Determine the proportion to be trimmed from the lower end of the sample (γ_l) by the proportion

$$\gamma_l = \gamma \left[\frac{UW_x}{UW_x + LW_x} \right] \quad [12]$$

where UW_x and LW_x are respectively the portion of the numerator and denominator of the previously defined statistic (Q_1). The notation for UW_x and LW_x are as follows:

$$UW_{Q_1} = U_{0.2} - L_{0.2}$$

and

$$LW_{Q_1} = U_{0.5} - L_{0.5}$$

Subsequently, the calculation for HQ_1 is

$$HQ_1 = \frac{UW_{Q_1}}{UW_{Q_1} + LW_{Q_1}} \quad [13]$$

- iii. The upper trimming percentage is defined as:

$$\gamma_u = \gamma - \gamma_l$$

3. Empirical Investigation

This study investigated on the performance (robustness) of F_t statistic using different types of trimmed mean as the central tendency measure for the case of two unbalanced groups. In studying the robustness of the procedures, two

variables were manipulated, creating conditions which are known to highlight the strengths and weaknesses of the tests. The two variables were: (1) population distribution and (2) degree of variance heterogeneity.

Unequal group sizes, when paired with equal and unequal variances, can affect Type I error control for some statistical tests. Thus, several ratios of variances were considered, namely, 1:1, 1:8 and 1:36. The total sample size for the two groups was set at $N = 30$ (12, 18). To test for the effect of distributions, the g - and h - distribution (Hoaglin, Mosteller & Tukey, 1983) was used to represent the skewed and normal data. Three types of distributions representing normal, skewed normal-tailed and skewed leptokurtic were considered. Observations of the g - and h - distribution were generated by converting the standard normal variates using the following equation:

$$Y_j = \begin{cases} \frac{\exp(gZ_j) - 1}{g} \exp(hZ_j^2 / 2) & \text{for } g \neq 0 \\ Z_j \exp(hZ_j^2 / 2) & \text{for } g = 0 \end{cases} \quad [14]$$

The g - and h - distribution is modified from standard 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 normal 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, 2005).

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. 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 $\alpha = 0.05$.

4. Results

The performance of the nine F_t procedures under unequal sample sizes with various variances is shown in Table 2. By convention, a procedure can be considered robust if its Type I error rates is between 0.5α to 1.5α (Bradley, 1978). Thus, when the nominal level is set at $\alpha = 0.05$, the Type I error rate should be in between 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.

4.1 Normal distribution

For normal distribution with $g = 0$ and $h = 0$, under $N = 30$, the F_t procedure using fixed trimmed mean with 15% and 20% symmetric trimming produced Type I error rates within robust criterion regardless of the variances. In contrast, fixed asymmetric trimming was able to produce good control of Type I error rates for 10% trimming under all variance ratios. For automatic trimming, F_t with $MOM - T_n$ and $MOM - LMS_n$ produced robust procedure when the ratio of variances were 1:8 and 1:36.

4.2 Skewed normal-tailed distribution

F_t with fixed symmetric trimmed mean produced all robust result except for 10% trimming with variance ratio of 1:36. As for automatic trimming, only F_t with $MOM - T_n$ having variance ratio of 1:8 and 1:36, and F_t with $MOM - LMS_n$ having 1:8 variance ratio were able to show Type I error within robust criterion. Nevertheless, fixed asymmetric trimming produced good control of Type I error for all variances ratio except when heterogeneity was 1:36 for 20% trimming.

4.3 Skewed leptokurtic distribution

When using equal variance 1:1, all Type I error rates for fixed symmetric, automatic and fixed asymmetric trimming produced Type I error rates within the Bradley's criterion of robustness except for fixed asymmetric with 15% and 20% trimming. Under unequal variances, the procedures with 15% and 20% symmetric trimming, $MOM - MAD_n$ and 15% asymmetric trimming produced robust procedure.

5. Discussion and conclusions

To evaluate the robustness of a test, several other benchmarks have been used in the past. Procedures that were considered not robust for some researchers could be deemed as robust for others. Some researchers would consider that the procedures with conservative Type I error rates fail to perform. However, Mehta and Srinivasan (1970) and Hayes (2005) stated that conservative procedures in which the true Type I error rate is less than or equal to the nominal level can still be considered as robust. Yet, a conservative test will be lower in power than a less conservative test because a more conservative test is less likely to reject any null hypothesis (Hayes, 2005). While for liberal tests, Hayes (2005) also had a piece of advice, i.e. to avoid using such tests. He defined liberal test as a test that tends to underestimate the true p -value. Using a liberal test for testing hypothesis will increase the probability of Type I error to a value greater than the nominal level, which implies that there is a bigger risk of making a Type I error. Nonetheless, Keselman *et al.*, (2000) pointed out that there is no one universal standard by

which tests can be judged to be robust, so different interpretations of these results are possible. This study also identified some promising procedures that performed well in terms of Type I error.

Based on the results, for F_t with fixed trimmed mean, 15% symmetric trimming is the best alternative because this procedure was able to produce Type I error rates which were very close to the nominal level for all the types of distributions tested (normal, skewed normal tailed and skewed leptokurtic). For F_t with automatic trimming, the $MOM - MAD_n$, $MOM - T_n$ and $MOM - LMS_n$ performed well for most of the conditions under skewed distribution regardless of variance ratios. As for F_t with asymmetric trimming using hinge estimator, 10% trimming showed good result for normal distribution while 15% trimming produced good control of Type I error rates for all variance ratios under skewed normal-tailed distribution which the Type I error rates are the nearest to the nominal level compared than the other trimming percentages.

Overall, the best procedure that should be taken into consideration is the F_t with $MOM - T_n$ for normal distribution, 15% fixed symmetric trimming for skewed normal-tailed distribution and $MOM - MAD_n$ for skewed leptokurtic distribution. All of these procedures produced the nearest Type I error rates to the nominal level.

References

- 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 - 152.
- Grubel, R. (1988). The length of the shorth. *The Annals of Statistics*, 16, 619 – 628, doi:10.1214/aos/1176350823, <http://dx.doi.org/10.1214/aos/1176350823>
- Hayes, A. F. (2005). *Statistical methods for communication science*. Mahwah, NJ: Erlbaum.
- Hoaglin, D. C., Mosteller, F., & Tukey, J. W. (1983). *Understanding robust and exploratory data analysis*. New York: Wiley.
- James, G. S. (1951). The comparison of several groups of observations when the ratios of the population variances are unknown. *Biometrika*, 38, 324-329.
- Keselman, H. J., Wilcox, R. R., Taylor, J., & Kowalchuk, R. K. (2000). Tests for mean equality that do not require homogeneity of variances: Do they really work? *Communications in Statistics*, 29, 875-895, doi:10.1080/03610910008813644, <http://dx.doi.org/10.1080/03610910008813644>
- Keselman, H. J., Kowalchuk, R. K., Algina, J., Lix, L. M., & Wilcox, R. R. (2000). Testing treatment effects in repeated measure designs: Trimmed means and bootstrapping. *British Journal of Mathematical and Statistical Psychology*, 53, 175 – 191.
- Keselman, H. J., Wilcox, R. R., Othman, A. R., & Fradette, K. (2002). Trimming, transforming statistics, and bootstrapping: Circumventing the biasing effects of heteroscedasticity and nonnormality. *Journal of Modern Applied Statistical Methods*, 1, 288 – 309.
- Keselman, H. J., Wilcox, R. R., Lix, L. M., Algina, J., & Fradette, K. H. (2007). Adaptive robust estimation and testing. *British Journal of Mathematical and Statistical Psychology*, 60, 267 - 293.
- 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.
- Lix, L. M., & Keselman, H. J. (1998). To trim or not to trim: Tests of location equality under heteroscedasticity and nonnormality. *Educational and Psychological Measurement*, 58(3), 409 – 429.
- Locker, L. A. (2001). *Some "modern" statistics: A primer and demonstration* (Annual Meeting of the Southwest Educational Research Association (New Orleans, LA)).
- Luh, W.-M. (1999). Developing trimmed mean test statistics for two-way fixed-effects ANOVA models under variance heterogeneity and nonnormality. *Journal of Experimental Education*, 67(3), 243-265.
- Luh, W.-M., & Guo, J.-H. (2005). Heteroscedastic test statistics for one-way analysis of variance: The trimmed means and Hall's transformation conjunction. *The Journal of Experimental Education*, 74(1), 75-100, doi:10.3200/JEXE.74.1.75-100, <http://dx.doi.org/10.3200/JEXE.74.1.75-100>
- Mehta, J. S., & Srinivasan, R. (1970). On the Behrens-Fisher problem. *Biometrika*, 57, 649-655, doi:10.1093/biomet/57.3.649, <http://dx.doi.org/10.1093/biomet/57.3.649>
- Myers, L. (1998). Comparability of the James' second-order approximation test and the Alexander and Govern \underline{A} statistic for non-normal heteroscedastic data. *Journal of Statistical Computational Simulation*, 60, 207-222.
- Othman, A. R., Keselman, H. J., Padmanabhan, A. R., Wilcox, R. R., & Fredette, K. (2004). Comparing measures of the 'typical' score across treatment groups. *British Journal of Mathematical and Statistical Psychology*, 57, 215-234.

Reed III, J. F., & Stark, D. B. (1996). Hinge estimators of location: Robust to asymmetry. *Computer Methods and Programs in Biomedicine*, 49, 11 – 17.

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

Rousseeuw, P. J., & Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, 88, 1273 – 1283.

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

Schneider, P. J., & Penfield, D. A. (1997). Alexander and Govern's approximation: Providing an alternative to ANOVA under variance heterogeneity. *Journal of Experimental Education*, 65, 271-287, doi:10.1080/00220973.1997.9943459, http://dx.doi.org/10.1080/00220973.1997.9943459

Welch, B. L. (1951). On the comparison of several mean values: An alternative approach. *Biometrika*, 38, 330-336, doi:10.1093/biomet/38.3-4.330, http://dx.doi.org/10.1093/biomet/38.3-4.330

Wilcox, R. R. (2002). Understanding the practical advantages of modern ANOVA methods. *Journal of Clinical Child and Adolescent Psychology*, 31(3), 399-412, doi:10.1207/153744202760082667, http://dx.doi.org/10.1207/153744202760082667

Wilcox, R. R. (2003). Multiple comparisons based on a modified one-step M-estimator. *Journal of Applied Statistics*, 30(10), 1231-1241, doi:10.1080/0266476032000137463, http://dx.doi.org/10.1080/0266476032000137463

Wilcox, R. R. (2005). *Introduction to robust estimation and hypothesis testing (2nd ed)*. San Diego, CA: Academic Press.

Wilcox, R. R., & Keselman, H. J. (2003). Modern robust data analysis methods: Measures of central tendency. *Psychological Methods*, 8, 254-274, doi:10.1037/1082-989X.8.3.254, http://dx.doi.org/10.1037/1082-989X.8.3.254

Table 1. Some properties of the g-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 leptokurtic

Table 2. Comparisons of Type I Error Rates

Dist.	Var.	Trimming methods								
		Fixed symmetric			Automatic			Fixed asymmetric		
		10%	15%	20%	<i>MAD_n</i>	<i>T_n</i>	<i>LMS_n</i>	10%	15%	20%
<i>g</i> = 0.0 <i>h</i> = 0.0	1:1	0.0476	0.0472	0.0478	0.0912	0.0810	0.1134	0.0498	0.0380	0.0244
	1:8	0.0220	0.0318	0.0274	0.0916	0.0508	0.0346	0.0284	0.0166	0.0116
	1:36	0.0208	0.0304	0.0276	0.0570	0.0492	0.0750	0.0310	0.0166	0.0116
Average		0.0301	0.0365	0.0343	0.0799	0.0603	0.0743	0.0364	0.0237	0.0159
<i>g</i> = 0.5 <i>h</i> = 0.0	1:1	0.0442	0.0450	0.0424	0.1078	0.1060	0.1252	0.0448	0.0450	0.0354
	1:8	0.0254	0.0352	0.0296	0.0780	0.0740	0.0644	0.0250	0.0330	0.0260
	1:36	0.0246	0.0358	0.0304	0.0758	0.0720	0.2198	0.0262	0.0300	0.0230
Average		0.0314	0.0387	0.0341	0.0872	0.0840	0.1347	0.0320	0.0360	0.0281
<i>g</i> = 0.5 <i>h</i> = 0.5	1:1	0.0318	0.0370	0.0326	0.0518	0.0474	0.0510	0.0258	0.0194	0.0138
	1:8	0.0196	0.0278	0.0218	0.0266	0.0246	0.0992	0.0144	0.0168	0.0122
	1:36	0.0190	0.0266	0.0206	0.0274	0.0244	0.4814	0.0148	0.0718	0.0126
Average		0.0235	0.0305	0.0250	0.0353	0.0321	0.2105	0.0183	0.0360	0.0129