Applied Mathematics and Computational Intelligence
Volume 8, No.1, Dec 2019 [77-88]

Modification of ANOVA with Various Types of Means

Nur Farah Najeeha Najdi[1*] and Nor Aishah Ahad[1]

[1]*School of Quantitative Sciences, Universiti Utara Malaysia.*

ABSTRACT

In calculating mean equality for two or more groups, Analysis of Variance (ANOVA) is a popular method. Following the normality assumption, ANOVA is a robust test. A modification to ANOVA is proposed to test the equality of population means. The suggested research statistics are straightforward and are compatible with the generic ANOVA statistics whereby the classical mean is supplemented by other robust means such as geometric mean, harmonic mean, trimean and also trimmed mean. The performance of the modified ANOVAs is then compared in terms of the Type I error in the simulation study. The modified ANOVAs are then applied on real data. The performance of the modified ANOVAs is then compared with the classical ANOVA test in terms of Type I error rate. This innovation enhances the ability of modified ANOVAs to provide good control of Type I error rates. In general, the results were in favor of the modified ANOVAs especially $ANOVA_T$ and $ANOVA_{TM}$.

Keywords: ANOVA, Geometric Mean, Harmonic Mean, Trimean, Trimmed Mean, Type I Error.

1. INTRODUCTION

Analysis of variance (ANOVA) is one of the most used model which can be seen in many fields such as medicine, engineering, agriculture, education, psychology, sociology and biology to investigate the source of the variations. One-way ANOVA is based on assumptions that the normality of the observations and the homogeneity of group variances. If the assumptions of normality and homogeneity of variances are invalid and also outliers are present, classical ANOVA does not give accurate results. Therefore, test statistics based on robust methods should be used instead of the classical ANOVA.

The one-way ANOVA under the violation of assumptions has been studied extensively. To deal with non-normal data and/or heteroscedastic variances across groups, many alternatives such as Q, Welch, Brown-Forsythe and Modified Brown-Forsythe tests have been developed instead of classical ANOVA. Note that with the normality of the probability distributions and the constant variability, the probability distributions differ only with respect to their means. ANOVA is a very powerful test as long as all prerequisites are met. It is not necessary, nor is it usually possible, that an ANOVA model fit the data perfectly. ANOVA models are reasonably robust against certain types of departures from the model, such as the data not being exactly normally distributed (Kutner, Nachtsheim, Neter & Li, 2005).

Basically, classical parametric tests such as analysis of variance (ANOVA) and independent sample t-test are often used in testing the central tendency measure by researchers rather than other methods since the aforementioned methods provide a good control of Type I error and generally more powerful than other methods when all the assumptions are fulfilled (Wilcox & Keselman, 2010). Analysis of variance (ANOVA) has been stated to be robust to departures from population normality (Glass, Peckham & Sanders, 1972).

*Corresponding Author: *nfarahnajeeha@gmail.com*

The calculation of ANOVA involves the mean. It is also known as the arithmetic mean. The simplest and most classical mean values are the arithmetic, the geometric and the harmonic mean. All of the means are known for their inequalities (Xia, Xu & Qi, 1999). The arithmetic mean is the most commonly used average. It is generally referred to as the average or simply mean. The breakdown point is zero. The arithmetic mean or simply mean is defined as the value obtained by dividing the sum of values by their number or quantity. The formula given below is the basic formula that forms the definition of arithmetic mean and is used in case of ungrouped data where weights are not involved. The formula is as follows,

$$A = \frac{a_1 + a_2 + a_3 + \cdots + a_n}{n} \tag{1}$$

Meanwhile, the geometric mean is used when numbers are multiplied. It is a useful application with percentage increases or decreases. Examples, where this kind of calculation may be applicable, include any growth measurement (where we are dealing with a constant time for change measurement) or in a financial situation where we are concerned with changes as a function of month or year. The geometric mean is calculated as the *n*th root of the product of the *n* positive observations. The formula for the geometric mean is as follows,

$$G = \sqrt[n]{x_1 x_2 x_3 \dots x_n} \tag{2}$$

Alternatively, the harmonic mean is used with inverse relationships. For example, speed and time are inversely related. In general, it is useful for expressing average rates. The harmonic mean is less than or equal to the geometric mean. Equality occurs when all the numbers are equal (Weisstein, 2003). The harmonic mean is the reciprocal of the arithmetic mean of the reciprocals of the observed values (Norris, 2000). The formula for the harmonic mean is as follows,

$$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} \tag{3}$$

Other resistant measures include the trimean, which is a central trend measure based on the arithmetic average value of the first quartile, the third quartile, and the median counted twice (Behrens, 1997). In statistics, the trimean is also known as Tukey's trimean (Tukey, 1977). The Trimean's foundations were part of the teachings of Arthur Bowley and later popularized in his 1977 book by statistician John Tukey, who gave his name to a set of techniques called exploratory data analysis (Doyle & Chen, 2009). Unlike the sample mean, with a breakdown point of 25%, it is a statistically resistant L-estimator. This beneficial property was described as the trimean's advantage as a center measure. It combines the emphasis of the median on core values with the attention of the midhinge to the extremes (Weisberg, 1992). The formula for the trimean is as follows,

$$T = \frac{1}{4}(Q_1 + 2Q_2 + Q_3) \tag{4}$$

The classical point and interval estimators use the sample mean and standard deviation. If a graph of the data indicates that the classical assumptions are violated, then an alternative estimator should be considered. Robust estimators can be obtained by giving zero weight to some cases and applying classical methods to the remaining data. Bickel (1965) and (Stigler, 1973) consider trimmed means or sometimes called a truncated mean. Shorack (1974) and also Shorack and Wellner (1986) derive the asymptotic theory for a large class of robust procedures. Special cases include trimmed, Winsorized, metrically trimmed, and Huber type skipped means. Many location estimators can be presented in a unified way by ordering the values of the sample as $x_1 \leq x_2 \leq x_3 \leq \cdots \leq x_n$.

A trimmed mean is calculated by discarding a certain percentage of the lowest and the highest scores and then computing the mean of the remaining scores. The prevalent method of trimming is to remove outliers from each tail of the distribution of scores. Tukey (1948) introduced trimmed mean as a compromise to the two classical estimators mean and median, to achieve balance between outlier(s) tolerance and efficiency. In addition, the recommendation is to trim 20% from each tail (Rosenberger & Gasko, 1983; Wilcox, 1995). The x% trimmed mean has a breakdown point of x% for the chosen level of x.

For example, a mean trimmed 50% is computed by discarding the lower and higher 25% of the scores and taking the mean of the remaining scores. The median is the mean trimmed 100% and the arithmetic mean is the mean trimmed 0%. Therefore, in this study, we will consider the robust method, trimmed mean along with the other means mentioned above. Wilcox (2012) states that trimming means cannot fix every problem but do work remarkably well to adjust for problems of heteroscedasticity (non-equal variances) and non-normality. The α-trimmed mean is as follows

$$\bar{x}_{tj} = \frac{1}{n_j - \alpha_{1j} - \alpha_{2j}} \left[\sum_{i=\alpha_{1j}+1}^{n_j - \alpha_{2j}} x_{ij} \right] \tag{5}$$

Thus, the trimmed mean corresponds to the mean value of data samples where p highest and p lowest samples are removed. Application of trimming lowers the influence of extreme data values on the result of averaging.

Outliers can completely break down the results of the ANOVA test when not properly taken into account (Wilcox, 1990). Given this limitation of the ANOVA test, there is a need for ANOVA type tests that are robust. Such an approach using robust estimators provides better control of the probability of the Type I error for one-way ANOVA situations (Lix & Keselman, 1998).

Calculation of the ANOVA has been using the classical mean or known as the arithmetic mean. It is very sensitive to changes in data series, especially to outliers. It has a breakdown point of 0%, so it shows that it is highly affected by extreme values or known as outliers. Therefore, it motivates us to perform this study by modifying the arithmetic mean with other type of means in ANOVA.

As for the geometric mean, it is usually used to tackle continuous data series in which it is unable to accurately be reflected by the arithmetic mean or the classical mean that is used in ANOVA. The arithmetic mean cannot average the ratios and percentages properly. Therefore, the geometric mean should be used since it has already been mentioned that it is useful for calculation regarding the increase or decrease of percentages. Since there is quite a few of continuous data series in real data, so in term of that, geometric mean would be preferable.

Other than that, a reliable estimation of the location of the bulk of the observations is needed. Therefore, the trimmed mean can help in providing a better estimation than the classical mean. Outliers and asymmetry less affect the standard error of the trimmed mean than the classical mean. ANOVA test using trimmed means can have more power than the test using the classical mean (Rousselet, Pernet & Wilcox, 2017). Therefore, in this study, we will compare the performance of ANOVA with the modified ANOVAs in term of the Type I error rate.

Therefore, the main objective for this study is to propose modification on classical ANOVA with geometric, harmonic, trimean and trimmed mean. The sub-objectives are to measure and compare performance between classical ANOVA with modified ANOVAs.

2. METHODOLOGY

2.1 Design Specification

The Monte Carlo simulation study was performed using the SAS programming language. Pseudo-random number generators were invoked to obtain random variates from the normal distribution by using the SAS generator (SAS, 2006). Normal variates with mean, μ = 0 and σ = 1 were generated. Nominal alpha was set at α = 0.05. Figure 3.1 shows the flow of the simulation study processes. The modified ANOVAs are then applied on real data.

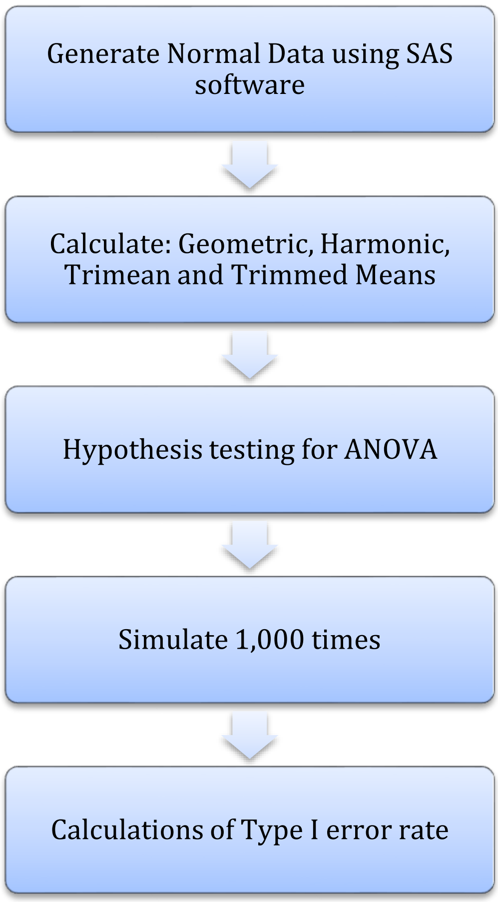

Figure 1. Flow chart of Simulation Study.

Each statistical method is assessed to compare three group means in a simulated dataset. We focus on three groups because it is quite common and can be found in a lot of trials. The sample size is to be manipulated in this simulation study such that small, medium and large sample size to test the robustness of the modified ANOVAs. We use group sizes of (15, 15, 15), (30, 30, 30) and (50, 50, 50) to represent the small, medium and large sample size, respectively. This is in order for us to compare the results from each size of group.

Type I error is the rejection of a true null hypothesis resulting in the faulty conclusion of statistically significant treatment effects. From the simulation, the number of count will increase by ‘1’ if the p-value is less than α=0.05 in which we reject the H_0. Then, the total number of counts were divided by the total simulations which is 1,000. Type I error rate corresponding to each method was determined and compared.

2.2 Robustness Criterion

In each simulation scenario, a test is considered to be significant when a p-value is less than the nominal α=0.05. The number of significant tests will be counted in simulated datasets in a scenario and the Type I error rate was calculated. The robustness of a method is determined by its ability in controlling the Type I error. Researchers have developed a few robustness criterions. Sullivan and D`Agostino in 2003 illustrated that a test which does not exceed 10% of the nominal significance level as robust. Guo and Luh (2000) interpreted that if a test's empirical Type I error rate is not higher than 0.075, with a 5% significance level, then it is robust.

Meanwhile, this study adopted the robustness liberal criterion by Bradley. This criterion was chosen since it was widely used by many robust statistic researchers (e.g. (Keselman, Kowalchuk, Algina, Lix, & Wilcox, 2000); (Othman, et al., 2004)) to judge robustness. A test can be considered robust if its empirical rate of Type I error, α, is within the interval 0.5α and 1.5α (Bradley, 1978). If the nominal level is α = 0.05, the empirical Type I error rate should be between 0.025 and 0.075.

2.3 ANOVA Test

The total variability can be split into several parts. The total amount of variability among observations can be measured by summing the squares of the differences between each x_{ij} and $\bar{x}$:

SST: Total sum of squares

$$\text{SST} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(x_{ij} - \bar{x}\right)^2 \tag{6}$$

The variability has two sources:

Variability between group means (specifically, variation around the overall mean $\bar{x}$)

$$\text{SSA} = \sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2, \text{ and} \tag{7}$$

Variability within groups means (specifically, variation of observations about their group mean $\bar{x}_i$)

$$\text{SSE} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(x_{ij} - \bar{x}_i\right)^2 = \sum_{i=1}^{k}(n_i - 1)s_i^2 \tag{8}$$

It is the case that,

$$\text{SST = SSA + SSE} \tag{9}$$

F-test is a measure of the variability between treatments divided by a measure of the variability within treatments. Table 1 shows the calculations for F-test for ANOVA.

Table 1 Table of ANOVA

Source	**SS**	**df**	**MS**	**F**
Model/ Group	SSA	k – 1	MSA = $\frac{SSA}{k-1}$	$\frac{MSA}{MSE}$
Residual/ Error	SSE	n – k	MSE = $\frac{SSE}{n-k}$	
Total	SST	n – 1		

2.4 Geometric Mean

The geometric mean is calculated as the nth root of the product of the n positive observations. The geometric mean of a population is given as,

$$\mathrm{G} = \sqrt[n]{x_1 x_2 x_3 \dots x_n} \tag{10}$$

Where $x_1, x_2, \dots, x_n$ are values of auxiliary variable of the population.

2.5 Harmonic Mean

Harmonic mean is another measure of central tendency and is also based on mathematics like arithmetic mean and geometric mean. Like arithmetic mean and geometric mean, harmonic mean is also useful for quantitative data. The harmonic mean is a very specific type of average. It is generally used when dealing with averages of units, like speed or other rates and ratios. Harmonic mean's formula is defined as:

$$\mathrm{H} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} \tag{11}$$

2.6 Trimean

In this paper, we consider another L–statistic, namely the sample trimean introduced by Tukey (1977) as an element of a set of statistical techniques in descriptive statistics called "exploratory data analysis". The trimean is defined as

$$\mathrm{T} = \frac{1}{4}(25th\ percentile + 2(median) + 75th\ percentile)$$

Or

$$\mathrm{T} = \frac{1}{4}(Q_1 + 2Q_2 + Q_3) \tag{12}$$

where Q_2 is the sample median and Q_1 and Q_3 are lower and upper hinges of the sample. The sample median is defined usually as

$$Q_2 = X_{\frac{n+1}{2}:n}, \qquad \text{if } n \text{ is odd,}$$
$$Q_2 = \frac{1}{2}\left(X_{\frac{n}{2}:n} + X_{\frac{n}{2}+1:n}\right), \qquad \text{if } n \text{ is even,}$$

which can be written in a more compact way as

$$Q_2 = \frac{1}{2}\left(X_{\left[\frac{n+1}{2}\right]:n} + X_{\left[\frac{n+1}{2}\right]:n}\right),$$

where $[x]$ and $[x]$ denote the floor and the ceiling functions defined as

$$[x] = \max\{n \in Z: n \le x\}, \qquad [x] = \min\{n \in Z: x \le n\},$$

2.7 Trimmed Mean

This study will trim the mean of the data by 20% as have been suggested by Rosenberger and Gasko (1983) and Wilcox (1995; 2005) that 20% trimming should be used. The sample trimmed mean is computed as shown below. Let $x_{1j}, x_{2j}, \ldots, x_{n_j j}$ be an ordered sample of group j with size n_j.

Calculate the α-trimmed mean of group j by using:

$$\bar{x}_{tj} = \frac{1}{h}\left[\sum_{i=g_1+1}^{n_j-g_2} x_{ij}\right] \tag{13}$$

Where

$$h = n_j - g_1 - g_2 \tag{14}$$
$$g_1 = [n_j \alpha_u] \tag{15}$$
$$g_2 = [n_j \alpha_1] \tag{16}$$

2.8 Modified ANOVA

This paper focuses on modifying the ANOVA with various type of means that are geometric mean, harmonic mean, trimean and trimmed mean. Their performances will be observed using the Type I error rate. As have been mentioned before in subsection 3.2, there are three parts in calculating the F value for ANOVA, which are SST, SSA and SSE.

Recalling the formula are as follows,

$$\mathrm{SST} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(x_{ij} - \bar{x}\right)^2$$

$$\mathrm{SSA} = \sum_{i=1}^{k} n_i\left(\bar{x}_i - \bar{x}\right)^2, \text{ and}$$

$$\mathrm{SSE} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(x_{ij} - \bar{x}_i\right)^2 = \sum_{i=1}^{k}(n_i - 1)s_i^2$$

SST = SSA + SSE

The $\bar{x}$ will be replaced by all the various type of means as shown in equations (10), (11), (12) and (13) to produce modified ANOVAs presented by $\mathrm{ANOVA_G}$, $\mathrm{ANOVA_H}$, $\mathrm{ANOVA_T}$ and $\mathrm{ANOVA_{TM}}$, respectively.

3. DATA ANALYSIS AND RESULTS

3.1 Simulation Results

The robustness of a method is determined by its ability to control the Type I error. Table 2 displays the empirical Type I error rates for all the procedures. Based on Bradley's liberal criterion of robustness (Bradley, (1978)), a test can be considered robust if the rate of Type I error, α is within the interval 0.5α and 1.5α. For the nominal level of α = 0.05, the Type I error rates should be between 0.025 and 0.075. 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

Methods	N=45(15,15,15)	N=90(30,30,30)	N=150(50,50,50)
ANOVA	0.0213	0.0351	0.0323
$ANOVA_G$	0.0270	0.0288	0.0281
$ANOVA_H$	0.0263	0.0280	0.0292
$ANOVA_T$	0.0354	0.0370	0.0397
$ANOVA_{TM}$	0.0366	0.0372	0.0403

All of the methods showed robust Type I error rates. The best procedure is $ANOVA_{TM}$ or modified ANOVA with trimmed mean which produces the nearest Type I error rate to the nominal level which is $\alpha = 0.05$.

3.2 Analysis on Real Data

The performance of the ANOVA and also the modified ANOVAs were demonstrated on real data. BMI of students was taken at random from secondary data. The data was collected from students in Carolinas Medical Center University, Charlotte, North Carolina, United States by Wayne W. LaMorte in 2006.

There are three groups of them. The sample sizes for Group 1, 2 and 3 were 20, 30 and 35 respectively. The results of the test in the form of p-values are given in Table 7. The Shapiro-Wilk test has been employed in order to determine the normality of data analysis.

Table 3 Tests of normality for BMI of group 1

	Shapiro-Wilk		
	Statistic	df	Sig.
BMI (Group 1)	.872	20	.537
a. Lilliefors Significance Correction			

Table 4 Tests of normality for BMI of group 2

	Shapiro-Wilk		
	Statistic	df	Sig.
BMI (Group 2)	.921	30	.619
a. Lilliefors Significance Correction			

Table 5 Tests of normality for BMI of group 3

	Shapiro-Wilk		
	Statistic	df	Sig.
BMI (Group 3)	.870	35	.597
a. Lilliefors Significance Correction			

Hypothesis:

H_0: The data is normally distributed.
H_1: The data is not normally distributed.
$\alpha = 0.05$

Normality Test for BMI of Group 1:

$p - value_{G1} = 0.537$
Since the $p - value_{G1} > \alpha$, failed to reject H_0.
Therefore, the data is normally distributed.

Normality Test for BMI of Group 2:

$p - value_{G2} = 0.619$
Since the $p - value_{G2} > \alpha$, failed to reject H_0.
Therefore, the data is normally distributed.

Normality Test for BMI of Group 3:

$p - value_{G3} = 0.597$
Since the $p - value_{G3} > \alpha$, failed to reject H_0.
Therefore, the data is normally distributed.

Based on the Shapiro-Wilk test as shown in Table 3, 4 and 5, all of the groups are normally distributed since the p-value is greater than α=0.05, which failed to reject the null hypothesis.

Table 6 Descriptive statistics for each group

Group	Sample size (N)	Mean of the BMI	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Min.	Max.
					Lower Bound	Upper Bound		
1	20	27.8	4.06	0.91	25.89	29.70	22	33
2	30	28.3	2.82	0.51	27.28	29.39	23	33
3	35	26.4	3.82	0.65	25.14	27.77	22	33

There are a total of 85 students from three different groups. In this data, the mean BMI of each group is 27.8kg/m^2, 28.3kg/m^2, 26.4kg/m^2 respectively. Furthermore, Group 2 has the lowest variation with value of 2.82 as compared to the other two groups.

Table 7 Results of the test using different methods

Methods	p-value
ANOVA	0.0330
$ANOVA_G$	0.0380
$ANOVA_H$	0.0380
$ANOVA_T$	0.0298
$ANOVA_{TM}$	0.0281

For comparison, the data were tested using all the five procedures mentioned in this study namely ANOVA and the modified ANOVAs, which are $ANOVA_G$, $ANOVA_H$, $ANOVA_{T,}$ and $ANOVA_{TM}$. As can be observed in Table 7, all of the methods show significant results in which they reject the null hypothesis. The $ANOVA_{TM}$ shows a better detection with the strongest significance (p = 0.0281) as compared to the other methods. As shown in the simulation results in Table 2, the $ANOVA_{TM}$ does produce robust Type I error rates. Even though $ANOVA_{TM}$ shows stronger significance (p = 0.0281) as compared to the other modified ANOVAs, but $ANOVA_{TM}$ in general only gave a brief information on the data since the data has been trimmed by 20%. Thus, misrepresentation of the result could occur.

4. CONCLUSIONS

The goals of this paper are to propose modification on classical ANOVA and also to measure and compare the performances in term of Type I error. The Type I error will increase when the method is less robust compared to other methods. This will cause wary rejection of the null hypothesis and the power of the test can be reduced. Undetected differences are one of the consequences. This study has integrated the ANOVA with various types of means such as geometric mean, harmonic mean, trimean and also trimmed mean.

This paper has shown some improvement in the statistical solution for detecting differences between location parameters. The findings showed that the modified robust procedures, $ANOVA_T$, $ANOVA_{TM}$, $ANOVA_G$, and $ANOVA_H$ are comparable with the classical ANOVA in controlling Type I error rates. In the analysis on real data, $ANOVA_T$ ($p = 0.0298$) and $ANOVA_{TM}$ ($p = 0.0281$) showed a slightly stronger significance than the classical ANOVA ($p = 0.0330$). $ANOVA_G$ ($p = 0.0380$) and $ANOVA_H$ ($p = 0.0380$) showed a weaker performance as compared to the classical ANOVA.

Even though the study has achieved its goals, some constraints were inevitable. Study limitations would be time restrictions since only 12 weeks are part of one semester. This implies that there is very little time to analyze more information. To improve the performance of the modified ANOVA methods, other types of mean or robust scale estimators should be considered to be replaced in the formula. There are plenty of them that can be chosen from.

ACKNOWLEDGMENTS

I am very grateful to Allah S.W.T for bestowing me the courage in completing this project successfully. I am delighted to convey my gratitude to all of the people who help in contributing in different ways in this project. It would not have been possible without the support and assistance from lecturers and experts.

REFERENCES

[1] Behrens, J. T. Principles and Procedures of Exploratory Data Analysis. Psychological Methods **2**, 2 (1997) 131-160.

[2] Bickel, P. J. On Some Robust Estimates of Location. Ann. Math. Stat. **36** (1965) 847-858.

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

[4] Doyle, J. R. &Chen, C. H. On the efficiency of the Trimean and Q123. Journal of Statistics and Management Systems **12**, 2 (2009) 319-323.

[5] Glass, G. V., Peckham, P. D. & Sanders, J. R. Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. Review of Educational Research **42** (1972) 237-288.

[6] Guo, J. H. & Luh, W. M. Statistics & Probability Letters **49** (2000) 1-7.

[7] Inc., S. I. SAS online doc. Cary, NC: SAS Institute Inc, (2006).

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

[9] Kutner, M. H., Nachtsheim, C. J., Neter, J. & Li, W. Applied Linear Statistical Models. Singapore City: McGraw-Hill, (2005).
[10] Lix, L. & Keselman, H. To trim or not to trim: Tests of location equality under heteroscedasticity. Educ. Psychol. Meas **58** (1998) 409-429.
[11] Norris, N. The standard errors of the geometric and harmonic means and their application to index numbers. Ann. Math. Statistics **11** (2000) 445-448.
[12] Othman, A. R., Keselman, H. j., Padmanabhan, A. R., Wilcox, R. R. & Fradette, K. Comparing measures of the 'typical' score across treatment groups. British Journal of Mathematical and Statistical Psychology **57**, 2 (2004) 215-234.
[13] Rosenberger, J. L. a& Gasko, M. Comparing location estimators: trimmed means, medians and trimean. Understanding Robust and Exploratory Data Analysis, (1983) 297-338.
[14] Rousselet, G., Pernet, C. & Wilcox, R. R. Beyond differences in means: robust graphical methods to compare two groups in neuroscience. figshare, (2017).
[15] Shorack, G. R. Random means. Ann. Statist. **1** (1974) 661-675.
[16] Shorack, G. R. & Wellner, J. A. Empirical Processes with Applications to Statistics. New York: Wiley, (1986).
[17] Stigler, S. M. The Asymptotic Distribution of the Trimmed Mean. Ann. Math. Statist., **1** (1973) 472-477.
[18] Sullivan, I. M. & D`Agostino, R. B. Statistics in Medicine **22** (2003) 1317-1334.
[19] Tukey, J. W. Approximate Weights. The Annals of Mathematical Statistics **19**, 1 (1948) 91-92.
[20] Tukey, J. W. Exploratory Data Analysis. Pearson, (1977).
[21] Weisberg, H. F. Central Tendency and Variability. Sage University, (1992).
[22] Weisstein, E. W. CRC Concise Encyclopedia of Mathematics. Chapman and Hall/CRC. Retrieved from Chapman and Hall/CRC, (2003).
[23] Wilcox, R. R. Comparing the means of two independent groups. Biom. J. **32** (1990) 771-780.
[24] Wilcox, R. R. ANOVA: A paradigm for low power and misleading measures of effect size? Review of Educational Research **65**, 1 (1995) 51-77.
[25] Wilcox, R. R. Trimming and Winsorization. Encyclopedia of Biostatistics, (2005).
[26] Wilcox, R. R. Modern Statistics for the Social and Behavioral Sciences: A Practical Introduction. CRC Press, (2012).
[27] Wilcox, R. R. & Keselman, H. J. Modern robust data analysis methods: Measures of central tendency. Psychological Methods **8** (2003) 254-274.
[28] Wilcox, R. R. & Keselman, H. J. Modern robust data analysis methods. Measures of central tendency, (2010) 1-43.
[29] Xia, D., Xu, S. & Qi, F. A Proof of the Arithmetic Mean-Geometric Mean-Harmonic Mean Inequalities. RGMIA Research Report Collection **2**, 1 (1999) 85-88.