

Malaysian Journal of Social Sciences and Humanities (MJSSH)

Volume 4, Issue 4, August 2019

e-ISSN : 2504-8562

Journal home page:
www.msocialsciences.com

Modification of ANOVA with Trimmed Mean

Nur Farah Najeeha Najdi[1], Nor Aishah Ahad[1]
[1]School of Quantitative Sciences, Universiti Utara Malaysia (UUM)

Correspondence: Nur Farah Najeeha Najdi (nfarahnajeeha@gmail.com)

Abstract

Analysis of Variance (ANOVA) is a well-known method to test the equality of mean for two or more groups. ANOVA is a robust test under the normality assumption. Arithmetic mean is used in the computation of the ANOVA test. Mean is known to be sensitive towards outlier and this problem will affect the robustness and power of ANOVA. In this study, modification of ANOVA was created using one type of mean to replace arithmetic mean namely trimmed mean. New approaches were be obtained for the computation of ANOVA. This study was conducted based on a simulation study and application on real data. The performance of the modified ANOVA is then compared with the classical ANOVA test in terms of Type I error rate. This innovation enhances the ability of modified ANOVA to provide good control of Type I error rates. The findings were in favor of the modified ANOVA or better known as $ANOVA_{TM}$.

Keywords: anova, modified anova, arithmetic mean, trimmed mean, type I error

Introduction

Analysis of variance (ANOVA) is one of the most frequently used methods in statistics (Moder, 2016). It allows us to compare more than two group means in a continuous response variable. An ANOVA model assumes that:

i. The probability distribution of responses in each group is normal.
ii. Each probability distribution has the same variance.
iii. Samples are independent.

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).

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:

$$A = \frac{a_1+a_2+a_3+\cdots+a_n}{n} \quad (1)$$

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:

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

Thus, the trimmed mean correspond 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).

Problem Statement

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 mean in ANOVA.

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 ANOVA in term of the Type I error rate.

Objectives of the Study

The followings are the objectives for this study:

1. To propose modification on classical ANOVA with trimmed mean.
2. To measure performance of the modified ANOVA in term of Type I error.
3. To compare performance between classical ANOVA with modified ANOVA.

Methodology

Design Specification

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

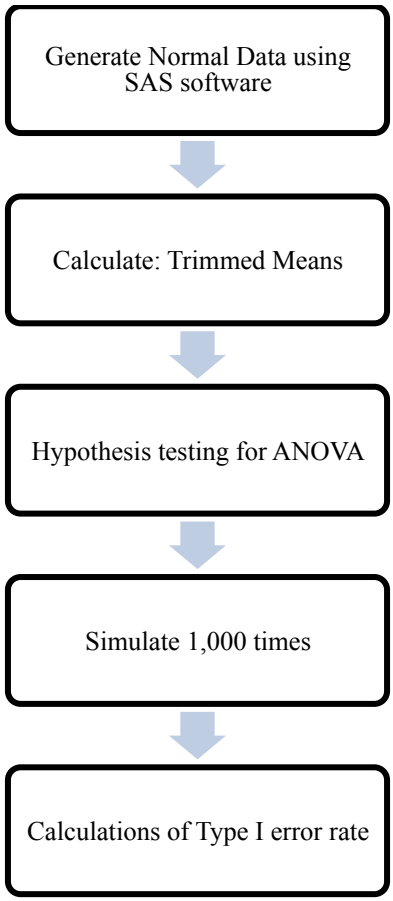


Figure 1: Flow chart of Simulation Study

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 ANOVA. Since this study uses ANOVA to compare three group means, the size of each group will be given as (a, b, c). 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.

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.

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.

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

$$\text{SST} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(x_{ij} - \bar{x}\right)^2 \quad (3)$$

The variability has two sources:

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

$$\text{SSA} = \sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2, \text{ and} \quad (4)$$

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

$$\text{SSE} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(x_{ij} - \bar{x}_i\right)^2 = \sum_{i=1}^{k}(n_i - 1)s_i^2 \quad (5)$$

It is the case that,

$$\text{SST} = \text{SSA} + \text{SSE} \quad (6)$$

F-test is a measure of the variability between treatments divided by a measure of the variability within treatments. Table 3.1 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		

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 $x_{1j}, x_{2j}, \dots, x_{n_j j}$ be an ordered sample of group j with size n_j.

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

$$\bar{x}_{tj} = \frac{1}{h}\left[\sum_{i=g_1+1}^{n_j-g_2} x_{ij}\right] \quad (7)$$

where

$$h = n_j - g_1 - g_2 \quad (8)$$

$$g_1 = [n_j \alpha_u] \quad (9)$$

$$g_2 = [n_j \alpha_1] \quad (10)$$

Modified ANOVA

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

Recalling the formula are as follows,

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

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

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

SST = SSA + SSE

The $\bar{x}$ will be replaced by the trimmed mean as shown in equation (7) to produce modified ANOVA presented by ANOVA_{TM}.

Data Analysis And Results

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
$ANOVA_{TM}$	0.0366	0.0372	0.0403

Both 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 $\alpha = 0.05$.

Analysis on Real Data

The performance of the ANOVA and also the modified ANOVA were demonstrated on real data. The data used in this study was the Index of African Governance. The Index of African Governance was a project of Harvard University's Kennedy School of Government's Program on Intrastate Conflict and Conflict Resolution and of the World Peace Foundation. This data set includes data on human development; participation and resource human; and safety and security of every Africa country.

There are three groups of them. The sample sizes for Group 1, 2 and 3 were 85, 53 and 32 respectively. 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.**
Human development (Group 1)	.872	85	.688

a. Lilliefors Significance Correction

Table 4: Tests of Normality for BMI of Group 2

	Shapiro-Wilk		
	Statistic	**df**	**Sig.**
Human development (Group 2)	.921	53	.752

a. Lilliefors Significance Correction

Table 5: Tests of Normality for BMI of Group 3

	Shapiro-Wilk		
	Statistic	**df**	**Sig.**
Human development (Group 3)	.870	32	.569

a. Lilliefors Significance Correction

Hypothesis:

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

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	Std. Deviation	Std. Error	Min.	Max.
1	85	64.5	2.74	0.81	31	91
2	53	62.9	2.93	0.48	30	82
3	32	74.0	3.75	0.47	31	88

There are a total of 85 students from three different groups. In this data, the mean of human development score of each group is 64.5, 62.9 and 74.0 respectively. They are quite the same in terms of average, thus showing that the respondents have about the same opinions thus same score regardless of how they are being grouped. Furthermore, Group 1 has the lowest variation with value of 2.74 as compared to the other two groups.

Table 7: Results of the test using different methods

Methods	p-value
ANOVA	0.0317
$ANOVA_{TM}$	0.0221

The results of the test in the form of p-values are given in Table 7. For comparison, the data were tested using all the five procedures mentioned in this study namely ANOVA and the modified ANOVA, which is $ANOVA_{TM}$. As can be observed in Table 7, both of the methods show significant results in which they reject the null hypothesis. The $ANOVA_{TM}$ shows a better detection with the strongest significance ($p = 0.0221$) as compared to the other method. 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 ($p = 0.0221$) as compared to the classical ANOVA, 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.

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 other type of mean, which is 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 procedure, $ANOVA_{TM}$ is comparable with the classical ANOVA in controlling Type I error rates. In the analysis on real data, $ANOVA_{TM}$ ($p = 0.0221$) showed a slightly stronger significance than the classical ANOVA ($p = 0.0317$).

Limitation of the Study

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.

Recommendation of the Study

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. For example, generalized means or power mean. It is an abstraction of the quadratic, arithmetic, geometric and harmonic means. Other example of type of mean that can be used is weighted arithmetic mean. Furthermore, it is recommended that in collecting data for any experiment, efforts should be made to always have large set of data to enable the central limit theorem to come into play should the normality assumption be violated. If the researcher want to use trimmed mean, then it is better to use a larger size of data sets. Other than that, type of data should be taken into account. This study has used growth data, therefore further research should use other data such as financial and economical data.

References

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

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

Bieniek, M. (2016). Sharp bounds on the bias of trimean. *57*, 365-379.

Bland, J. M. and Altman, D. G. (2007). Agreement between methods of measurement with multiple observations per individual. *Journal of Biopharmaceutical Statistics, 17*(4), 571-582.

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

De, P. (2016). The arithmetic mean - Geometric mean - Harmonic mean: Inequalities and a spectrum of applications. *Resonance, 21*(12), 1119-1133.

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

Evans. (1955). Appendix G. *Inefficient statistics*, 902-904. Retrieved from Inefficient statistics.

Ferger, W. F. (1931). The Nature and Use of the Harmonic Mean. *Journal of the American Statistical Association, 26*(173), 36-40.

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

Gleason, J. H. (2013). *Comparative Power of the ANOVA, Approximate Randomization ANOVA, and Kruskal-Wallis Test.*Wayne State University.

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

Hasan, M. K., Othman, N. A., Karim, S. A. and Sulaiman, J. (2018). Semi Non-Standard Trimean Algorithm for Rosenzweig-MacArthur Interaction Model. *International Journal on Advanced Science, Engineering and Information Technology, 8*(4-2), 1520.

Heeren, T. and D'Agostino, R. B. (1987). Statistics in Medicine. *6*, 79-90.

Huber, P. J. (2004). *Robust statistics.* New York: Wiley.

Hunter, M. A. and May, R. B. (1993). Some myths concerning parametric and nonparametric tests. *Canadian Psychology/Psychologie canadienne, 34*, 384-389.

Inc., S. I. (2006). *SAS online doc.* Cary, NC: SAS Institute Inc.

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.

Komic, J. (2011). Harmonic Mean. *International Encyclopedia of Statistical Science*, 622-624.

Kundu, S. (2013). *An Introduction to Business Statistics.*

Kutner, M.H., Nachtsheim, C.J., Neter, J. and Li, W. (2005). *Applied Linear Statistical Models.* Singapore City: McGraw-Hill.

Li, N. (2015). Estimating Life Tables for Developing Countries. *Technical Paper No. 2014/4*, pp. 1-29.

Lix, L. and Keselman, H. (1998). To trim or not to trim: Tests of location equality under heteroscedasticity. *Educ. Psychol. Meas, 58*, 409-429.

Luh, W. M. and Guo, J. H. (2010). Approximate Sample Size Formulas for Testing Group Mean Differences When Variances Are Unequal in One-Way ANOVA. *Educational and Psychological Measurement, 68*(6), 959.

Mendes, M. and Yigit, S. (2012). Comparison of ANOVA-Fand ANOM tests with regard to type I error rate and test power. *Journal of Statistical Computation and Simulation, 83*(11), 2093-2104.

Moder, K. (2016). How to keep the Type I Error Rate in ANOVA if Variances are Heteroscedastic. *Austrian Journal of Statistics, 36*(3), 179.

Mosteller, F. (2006). On some useful "inefficient" statistics. *Hoaglin DG (ed) Selected Papers of Frederick Mosteller*, 69-100.

Mu, W. and Yuan, X. (2012). Statistical inference for ANOVA under heteroscedasticity: Statistical inference. *2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet)*.

Norris, N. (2000). The standard errors of the geometric and harmonic means and their application to index numbers. *Ann. Math. Statistics, 11*, 445-448.

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

Permutt, T. and Li, F. (2017). Trimmed means for symptom trials with dropouts. *Pharm. Stat., 16*(1), 21.

Rhoad, R., Milauskas, G. and Whipple, R. (1991). *Geometry for enjoyment and challenge.* Evanston, IL: McDougal Littell.

Rocke, D. M., Downs, G. W. and Rocke, A. J. (1982). Are Robust Estimators Really Necessary? *Technometrics, 24*(2), 95-101.

Rosenberger, J. L. and Gasko, M. (1983). Comparing location estimators: trimmed means, medians and trimean. *Understanding Robust and Exploratory Data Analysis*, 297-338.

Rousselet, G., Pernet, C. and Wilcox, R. R. (2017). *Beyond differences in means: robust graphical methods to compare two groups in neuroscience.* figshare.

Santos, R. G., Giulianotti, M. A., Dooley, C. T., Pinilla, C., Appel, J. R. and Houghten, R. A. (2011). The Use and Implications of the Harmonic Mean Model on Mixtures for Basic Research and Drug Discovery. *ACS combinatorial science, 13*(3), 337-344.

Sarkar, J. and Rashid, M. (2016). A geometric view of the mean of a set of numbers. *Teaching Statistics, 38*(3), 77-82.

Shorack, G. R. (1974). Random means. *Ann. Statist., 1*, 661-675.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics.* New York: Wiley.

Stigler, S. M. (1973). The Asymptotic Distribution of the Trimmed Mean. *Ann. Math. Statist., 1*, 472-477.

Sullivan, I. M. and D`Agostino, R. B. (2003). *Statistics in Medicine, 22*, 1317-1334.

Sullivan, I. M. and D'Agostino, R. B. (1996). Statistics in Medicine. *15*, 477-496.

Thelwall, M. (2015). The precision of the arithmetic mean, geometric mean and percentiles for citation data: An experimental simulation modelling approach. *Journal of Informetrics*.

Tiku, M. L., Tan, W. Y. and Balakrishnan, N. (1986). *Robust inference.* New York: Marcel Dekker.

Tukey, J. W. (1948). Approximate Weights. *The Annals of Mathematical Statistics, 19*(1), 91-92.

Tukey, J. W. (1977). *Exploratory Data Analysis.* Pearson.

Tukey, J. W. (1982). *Teaching of Statistics and Statistical Consulting.* Ft. Belvoir: Defense Technical Information Center.

Velina, M., Valeinis, J., Greco, L. and Luta, G. (2016). Empirical Likelihood-Based ANOVA for Trimmed Means. *International journal of environmental research and public health, 13*(10), 953.

Weisberg, H. F. (1992). *Central Tendency and Variability.* Sage University.

Weisstein, E. W. (2003). *CRC Concise Encyclopedia of Mathematics.* Chapman and Hall/CRC. Retrieved from Chapman and Hall/CRC.

Wilcox, R. R. (1990). Comparing the means of two independent groups. *Biom. J., 32*, 771-780.

Wilcox, R. R. (1995). ANOVA: A paradigm for low power and misleading measures of effect size? *Review of Educational Research, 65*(1), 51-77.

Wilcox, R. R. (2005). Trimming and Winsorization. *Encyclopedia of Biostatistics*.

Wilcox, R. R. (2012). *Modern Statistics for the Social and Behavioral Sciences: A Practical Introduction.* CRC Press.

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

Wilcox, R. R. and Keselman, H. J. (2010). Modern robust data analysis methods. *Measures of central tendency*, 1-43.

Xia, D., Xu, S. and Qi, F. (1999). A Proof of the Arithmetic Mean-Geometric Mean-Harmonic Mean Inequalities. *RGMIA Research Report Collection, 2*(1), 85-88.

Zitt, M. (2012). The journal impact factor: angel, devil, or scapegoat? A comment on JK. *Scientometrics, 92*(2), 485-503.