International Journal of Innovative Technology and Interdisciplinary Sciences
https://journals.tultech.eu/index.php/ijitis
ISSN: 2613-7305
Volume 8, Issue 2
DOI: https://doi.org/10.15157/IJITIS.2025.8.2.324-335
Received: 26.05.2025; Revised: 02.07.2025; Accepted: 06.07.2025

Research Article

A Trimean and Asymmetry-Based Statistical Permutation Test for Group Comparisons

Necati Alp Erilli[1*]

[1]Department of Econometrics, Sivas Cumhuriyet University, Sivas, Türkiye
***aerilli@cumhuriyet.edu.tr**

Abstract

Statistical multiple comparison tests are methods used to detect differences between several groups and to assess whether these differences are statistically significant. Typically, parametric tests like ANOVA (Analysis of Variance) or non-parametric tests like Kruskal-Wallis are employed for this purpose. In this study, we propose a new statistical permutation test based on the trimean and Bowley's measure of asymmetry as an alternative to conventional multiple comparison tests. The proposed method is compared with ANOVA and Kruskal-Wallis tests in terms of reliability and statistical power. The analyses demonstrate that the proposed test yields statistically significant and effective results comparable to traditional methods. The findings reveal that the new test provides reliable outcomes especially for heterogeneous groups, skewed distributions, and small sample sizes. Overall, the proposed method can be considered a viable alternative in statistical analysis.

Keywords: Multiple Comparison Tests; Trimean; Asymmetry Measure; Permutation.

INTRODUCTION

In statistical analyses, various tests are used to determine whether there are overall differences between the means or medians of multiple groups. If the data satisfy the assumption of normality and homogeneity of variance, ANOVA is the most commonly used test. When these assumptions are violated, non-parametric alternatives like the Kruskal-Wallis test are preferred. The choice of test may vary depending on repeated measurements and homogeneity of variances [1-3]. When primary tests indicate an overall difference, post-hoc (multiple comparison) tests are used to identify which specific groups differ. For normally distributed data, Tukey HSD, Scheffé, Dunnett, Games-Howell, Bonferroni, and Tamhane's T2 tests are commonly used, while for non-normally distributed data, Dunn and Holm-adjusted Mann-Whitney U tests are preferred [4].

Although ANOVA and Kruskal-Wallis are widely used and effective in many situations, they may be limited under assumption violations, complex research questions, or different data structures. ANOVA, while a robust and popular test, has certain limitations due to its reliance on parametric assumptions as follows [5]:

- *Normality Assumption:* The dependent variable is expected to follow a normal distribution in each group. Although the Central Limit Theorem helps in large samples, deviations in small samples can affect power and Type I error rates.
- *Homogeneity of Variance:* ANOVA assumes equal variances across groups. When variances are heterogeneous, the results may be unreliable.
- *Independence of Observations:* If observations are not independent, the ANOVA assumption is violated, and repeated measures ANOVA or mixed models should be used.

Furthermore, ANOVA only tests for overall differences and does not control for covariates directly. It also requires continuous dependent variables. The Kruskal-Wallis test, although non-parametric, also has limitations: it ignores distribution shape, cannot test for interaction effects between multiple factors, and like ANOVA, only identifies whether any difference exists across groups [6, 7].

When data are skewed or contain outliers, traditional methods like ANOVA and Kruskal-Wallis can yield misleading results. To address these issues, this study proposes a non-parametric permutation test that evaluates differences in both central tendency (trimean) and distribution shape (asymmetry). Moreover, both ANOVA and Kruskal-Wallis lack sensitivity to the distributional shape of the data. While Kruskal-Wallis is non-parametric, it assumes equal shape across groups, and this assumption can be problematic in real-world data where group distributions are heterogeneous in skewness or variance. Recent studies [8-9] emphasize that traditional tests may misrepresent group differences when the data structure deviates from normality or symmetry.

A LITERATURE REVIEW

In recent years, there has been a growing interest in finding alternatives to classical multiple comparison tests (ANOVA, Kruskal-Wallis, etc.) in the literature. Although classical methods are widely used in comparisons, they have some significant limitations. These shortcomings have motivated the development of alternative methods. Perhaps the most important of these is that the methods are overly dependent on assumptions. ANOVA and t-tests require normal distribution and homogeneity of variance. Real-world data often do not meet these assumptions. Violations of normality can lead to Type I or Type II errors. Non-parametric tests like Kruskal-Wallis do not require normality, but they may suffer from power loss and information loss in ordered data. Classical tests (especially parametric methods) are seriously affected by outliers. Similarly, methods such as classical ANOVA may experience overfitting or statistical power loss in high-dimensional data [8-10]. The primary objectives of this search for alternative methods, which is the subject of this research, can be summarised as follows: developing more robust methods independent of assumptions, better controlling multiple comparison errors, enabling more comfortable work with high-dimensional data, and obtaining more flexible and interpretable results.

Authors at [11] emphasised that classical ANOVA may be inadequate when its assumptions (particularly normality and variance homogeneity) are violated and showed that using multiple contrast tests can provide advantages in terms of both statistical power and interpretation. They noted that this approach is more flexible, especially for non-parametric and heterogeneous data. Authors in [12] introduced the 'Analysis of Means' (ANOM) method in a broader context, demonstrating how it can be applied under different distribution assumptions and in complex designs. [13] criticised the two-step approach of first performing a significance test with ANOVA and then applying secondary comparisons such as Dunnett. While arguing that this strategy is not appropriate either theoretically or practically and that better alternatives (e.g., multiple comparison tests performed in a single step) exist, he noted that the F-test prerequisite could be misleading. [14] provides a comprehensive overview of the history, fundamental principles, and applications of the ANOM (Analysis of Means) approach in his review article. [15] examined the applicability of numerical calculations and algorithms for multiple comparison procedures (particularly stepwise methods and multiple hypothesis tests) in software such as R. [16] developed non-parametric multiple comparison tests for two-sided but unbalanced (different group sizes) experimental designs. They noted that this method, proposed for situations where the assumptions of classical parametric methods are not met, provides more flexible and reliable results in practice.

Researchers at [17] proposed a new and improved statistical method for comparing group means. They tested the proposed method using both simulations and real data and noted that it offers lower error rates, high statistical power, and broader application areas. Authors in [18] proposed a ranked Multiple Contrast Test Procedure (MCTP) for use in situations involving distribution assumption violations, such as psychological data, and demonstrated that the method provides log odds-like effect sizes. Authors at [19] have provided robust alternatives to outliers and skewed data in R, extending one-way ANOVA with trimmed mean, quantile ANOVA, and robust post hoc tests. In [20] is proposed a robust Wald-type test for data with a log-normal distribution assumption and used an approach that is insensitive to outliers. Authors at [21] highlighted situations where classical parametric methods (e.g., Dunnett test) are sensitive in multiple comparisons against a control group and proposed a robust test approach that can provide reliable results under such conditions. They demonstrated that permutation tests and rank-based methods are more reliable in fields such as toxicology, where biological data often do not meet the normality assumption. Recently, [9] proposed a unified framework for robust group comparisons using quantile-based effect sizes and trimmed means. These techniques are particularly effective in data with outliers or heavy-tailed distributions. Although our proposed TABS method also targets robust group comparisons by incorporating both central tendency and asymmetry, a formal comparative analysis with such robust frameworks remains a valuable area for future investigation.

PROPOSED METHOD

The proposed permutation test for group comparisons based on trimean and quartile asymmetry (TABS) aims to provide reliable results in settings with skewed distributions and heterogeneous variances. It is based on differences in robust central tendency (trimean) and distribution shape (Bowley asymmetry coefficient).

The algorithm of the method will be as given below:

1. *Calculate the specified statistics for each group*

i. Trimean parameter, see equation (1):

$$Trimean = \frac{Q_1 + 2 \times Q_2 + Q_3}{4} \tag{1}$$

ii. Inter-Quartile Range, see equation (2):

$$IQR = Q_3 - Q_1 \tag{2}$$

iii. Bowley Asymmetry Measure, see equation (3):

$$AS_{Bowley} = \frac{Q_3 + Q_1 - 2 \times Q_2}{Q_3 - Q_1} \tag{3}$$

iv. Standard error of the Bowley asymmetry measure:

It is calculated by the bootstrap method. The ASB is calculated by taking 1000 replicates of the data, the standard deviation of these values is $SE(AS_B)$.

2. Calculate TABS statistics for pairwise differences between groups

Calculate values for all pairs of groups (e.g. A vs B, A vs C, B vs C...) with the TABS formula given in equation (4):

$$T_{ij} = \frac{\left|Trimean_i - Trimean_j\right|}{\sqrt{IQR_i^2 + IQR_j^2}} + \frac{\left|(AS_B)_i - (AS_B)_j\right|}{\sqrt{SE(AS_B)_i^2 + SE(AS_B)_j^2}} \tag{4}$$

Calculate this value for each pair and add them together, see equation (5):

$$T = \sum_{i<j} T_{ij} \tag{5}$$

3. TABS Test Calculation of statistics required for permutation p-value calculation

To determine the significance level, the group labels are randomly shuffled to create K permutation samples, each of which distorts the structure of the real data. For each permutation, the observed test statistic is calculated. To do this, first the group labels are randomly assigned and the $T_{perm}^{(k)}$ statistic defined below is recalculated, see equation (6).

$$T_{perm}^{(k)} = \sum_{i<j} \left(\frac{\left| Trimean_i^{(k)} - Trimean_j^{(k)} \right|}{\sqrt{IQR_i^{(k)2} + IQR_j^{(k)2}}} + \frac{\left| (AS_B^{(k)})_i - (AS_B^{(k)})_i \right|}{\sqrt{SE(AS_B^{(k)})_i^2 + SE(AS_B^{(k)})_{ij}^{22}}} \right) \quad (6)$$

This process is repeated K times and the permutation distribution is generated, see equation (7):

$$T_{perm}^{(1)}, T_{perm}^{(2)}, ..., T_{perm}^{(K)} \quad (7)$$

4. Finding the p-value based on permutation

The permutation-based p-value is calculated based on the frequency with which the observed test statistic is greater than (or equal to) the values in the permutation distribution:

$$p = \frac{1}{K} \sum_{k=1}^{K} I\left(T_{perm}^{(k)} \geq T \right) \quad (8)$$

Here I(.) is the indicator function and takes the value 1 if the condition is true and 0 if it is false.

The TABS statistic can test for differences in trimean and skewness between multiple groups but cannot produce a p-value on its own. Since no parametric distribution assumption is made, a permutation test is used to determine the "significance limit" of the TABS value. The permutation test calculates the p-value by repeating the question "what would the TABS statistic be if the groups were randomly assigned?" K times and comparing the observed TABS statistic with this distribution.

The proposed method provides a robust and flexible way of assessing the significance of the test statistic without relying on distributional assumptions. Thanks to this approach, it is possible to assess whether there are significant differences between groups in terms of central tendency and distributional structure without any distributional assumptions.

APPLICATION

The proposed method is tested on different data sets, evaluated with ANOVA and Kruskal-Wallis multiple comparison tests and the results are interpreted. Since ANOVA and Kruskal-Wallis tests are based on different assumptions, comparisons were made with both classical methods to see the cases where the proposed TABS method gave similar or different decisions. In application, 6 real data sets (Iris, Iris with 5% outliers, Iris with 10% outliers, mtcars, PlantGrowth, chickwts, InsectSprays and warpbreaks) which can easily be obtained in R software and 5 simulation data sets, were used for multiple comparisons. Outliers of 5% and 10% in Iris dataset and 5%, 10% and 20% in Simulation-3, Simulation-4 and Simulation-5 datasets were created and included in the analysis. Analyses were performed in the R software program (version 4.5.0) with codes written by the author and some publicly available packages (dplyr [22], ggplot2 [23]) were used in the R library. Statistical confidence level was taken as 0.05 in all analyses. Box-plot graphs for all real-

time data sets are provided in the Figure 1 and simulation data sets in the Figure 2 to better illustrate the distribution of the data used in the study.

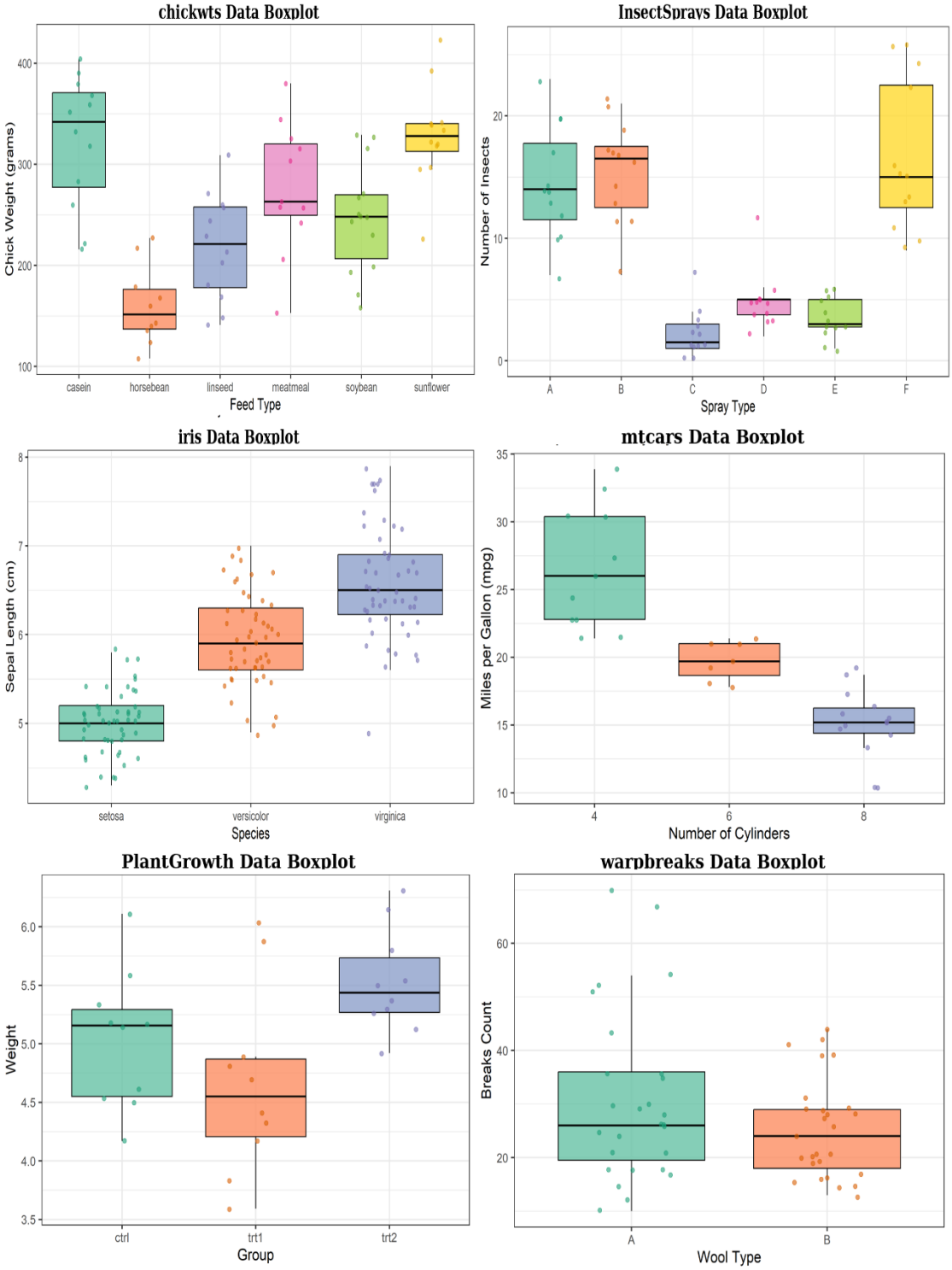

Figure 1. Boxplots for 6 real-time data sets

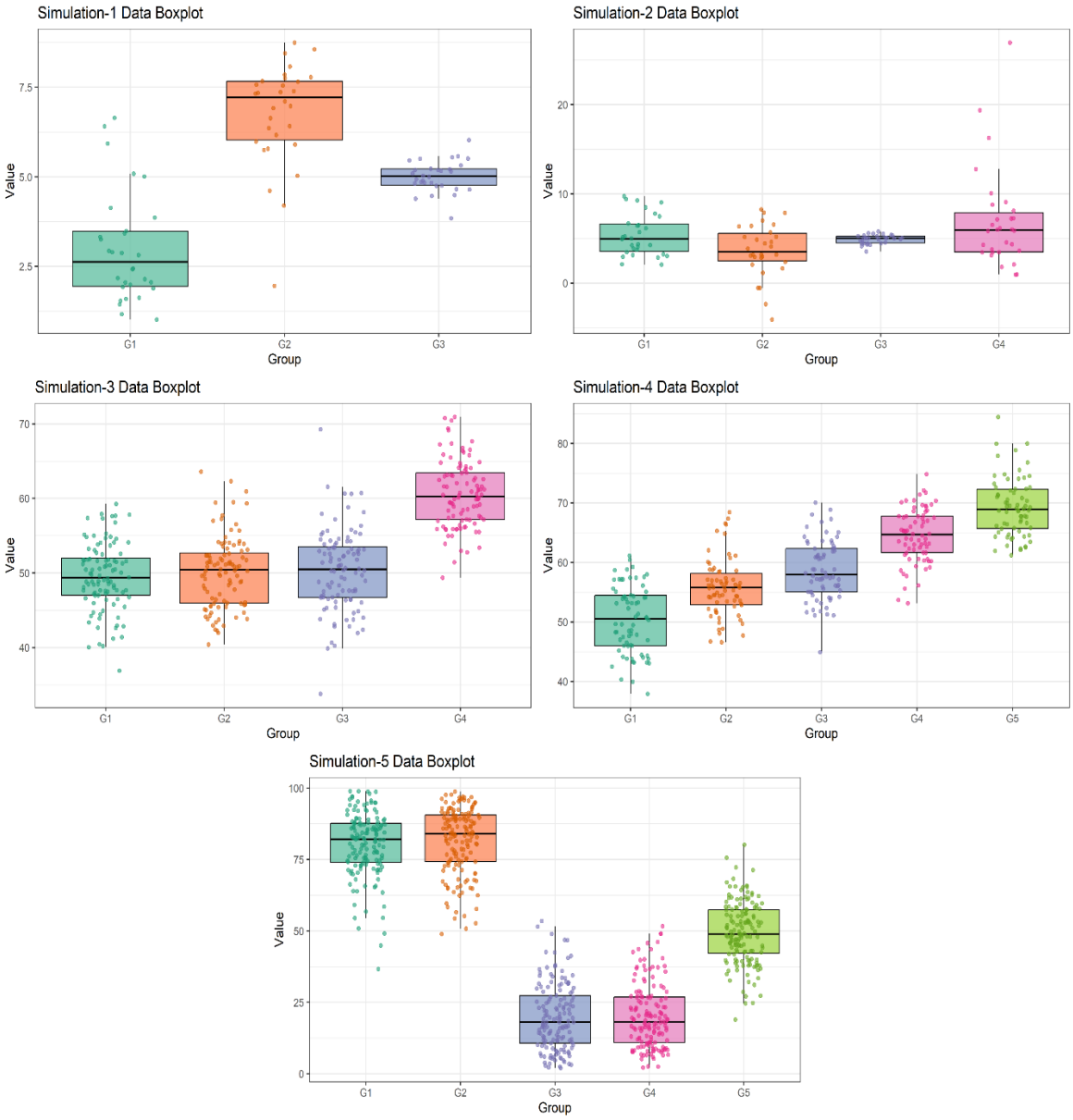

Figure 2. Boxplots for 5 simulation data sets

In Table 1, information about real-time data sets is given.

Table 1. Information for real-time data sets

Data Set	Sample Size	Number of Group	Skewness
iris	150	5	Some
mtcars	32	3	Some
PlantGrowth	30	2	Little
chickwts	71	2	High
InsectSprays	72	2	High
warpbreaks	54	3	High

First, 3 different multiple comparison methods mentioned above (ANOVA, Kruskal-Wallis, TABS) were applied to 8 different datasets (Iris, Iris with 5% outliers, Iris with 10% outliers, mtcars, PlantGrowth, chickwts, InsectSprays and warpbreaks). The obtained analysis results are given in Table 2.

Table 2. Results for real-time data sets

Method	Test Statistic	p-value	Data	Method	Test Statistic	p-value	Data
ANOVA	1180.16	$<10^{-90}$		ANOVA	4.846	0.0159	
K-Wallis	130.41	$<10^{-28}$	Iris	K-Wallis	7.988	0.01842	PlantGrowth
TABS	10,9	0.000		TABS	0.428	0.020	
ANOVA	58.37	$<10^{-26}$	Iris (with %5 outlier)	ANOVA	15.37	$<10^{-10}$	
K-Wallis	107.39	$<10^{-22}$		K-Wallis	37.343	$<10^{-7}$	chickwts
TABS	8.55	0,0067		TABS	46.36	0.000	
ANOVA	29.51	$<10^{-16}$	Iris (with %10 outlier)	ANOVA	34.7	$<10^{-16}$	
K-Wallis	88.26	$<10^{-19}$		K-Wallis	54.691	$<10^{-10}$	InsectSprays
TABS	9.36	0.008		TABS	8.108	0.000	
ANOVA	39.7	$<10^{-9}$		ANOVA	5.828	0.000277	
K-Wallis	25.75	$<10^{-6}$	mtcars	K-Wallis	15.778	0.0075	warpbreaks
TABS	3.87	0.000		TABS	93.74	0.000	

Five different simulation data sets were used in the study. In three of these, outliers were created at rates of 5%, 10% and 20% to test the strength of the proposed method in data structures containing outliers. The following R codes were used to obtain the simulation data: symmetrical data were generated using a symmetrical normal distribution (rnorm), left-skewed data were generated using a negative log-normal distribution (-1*rlnorm), and right-skewed data were generated using a log-normal distribution (rlnorm) or a gamma distribution (rgamma). The obtained groups were combined using the set.seed and data.frame codes to create the final simulation data. The floor and df$Value(df$Group) codes were also used to add outliers to the existing data sets. In Table 3, information about simulation data sets is given.

Table 3. Information for simulation data sets

Data Set	Sample Size	Number of Group	Status
Simulation-1	90	3	Each groups is Normal distributed
Simulation-2	120	4	2 right skewed group
Simulation-3	400	4	1 right skewed, 1 left skewed group
Simulation-4	350	5	2 right skewed, 1 left skewed group
Simulation-5	750	5	2 right skewed, 2 left skewed group

The test statistics and p-values obtained from five different simulation data sets for ANOVA, Kruskal-Wallis, and TABS methods are presented in Table 4.

Table 4. Results for Simulation data sets

Method	Test Sta.	p-value	Data
ANOVA	27.22	$<10^{-8}$	Sim-1
K-Wallis	32.66	$<10^{-6}$	
TABS	7.2	0.002	
ANOVA	4.97	0.0028	Sim-2
K-Wallis	8.42	0.038	
TABS	6.5	0.12	
ANOVA	121.4	$<10^{-55}$	Sim-3 (%0 Outlier)
K-Wallis	181.19	$<10^{-38}$	
TABS	9.83	0.000	
ANOVA	60.48	$<10^{-31}$	Sim-3 (%5 Outlier)
K-Wallis	132.52	$<10^{-27}$	
TABS	9.59	0.000	
ANOVA	40.19	$<10^{-21}$	Sim-3 (%10 Outlier)
K-Wallis	104.79	$<10^{-22}$	
TABS	8.64	0.04	
ANOVA	22.29	$<10^{-12}$	Sim-3 (%20 Outlier)
K-Wallis	74.27	$<10^{-15}$	
TABS	5.42	0.37	
ANOVA	152.11	$<10^{-74}$	Sim-4 (%0 Outlier)
K-Wallis	227.8	$<10^{-47}$	
TABS	14.31	0.000	
ANOVA	101.7	$<10^{-56}$	Sim-4 (%5 Outlier)
K-Wallis	211.56	$<10^{-43}$	
TABS	15.14	0.000	
ANOVA	57.82	$<10^{-37}$	Sim-4 (%10 Outlier)
K-Wallis	165.41	$<10^{-33}$	
TABS	12.54	0.03	
ANOVA	43.06	$<10^{-28}$	Sim-4 (%20 Outlier)
K-Wallis	131.42	$<10^{-27}$	
TABS	13.96	0.02	
ANOVA	1094.1	$<10^{-50}$	Sim-5 (%0 Outlier)
K-Wallis	616.9	$<10^{-50}$	
TABS	814.3	0.000	
ANOVA	262.63	$<10^{-50}$	Sim-5 (%5 Outlier)
K-Wallis	452.96	$<10^{-50}$	
TABS	705.6	0.000	
ANOVA	136.9	$<10^{-50}$	Sim-5 (%10 Outlier)
K-Wallis	329.3	$<10^{-50}$	
TABS	720.65	0.000	
ANOVA	51.575	$<10^{-38}$	Sim-5 (%20 Outlier)
K-Wallis	156.37	$<10^{-33}$	
TABS	528.4	0.000	

According to the results of the analysis, the results obtained according to different data sets can be briefly summarized as follows: In Iris data, TABS p-value increases as the outlier is added. In mtcars data, all tests yielded significant results. It is noteworthy that the TABS test is not significant in Simulation-2 data. In simulation-3 data, TABS p-value increases as the outlier increases, while classical tests always yielded significant results. In Simulation-4 data, TABS gave conservative p-values at 10% and 20% outliers.

DISCUSSION

The findings of this study reveal that classical statistical methods such as ANOVA and Kruskal-Wallis, although widely used, may produce overly optimistic results in the presence of skewness or outliers. In contrast, the proposed TABS (Trimean and Asymmetry-Based Statistical Permutation Test) demonstrate a flexible and adaptive behavior across various data conditions. Notably, in Simulation-3, which includes one right-skewed and one left-skewed group, the TABS test p-value increased as the proportion of outliers rose (from 0.000 at 0% to 0.37 at 20% outliers), whereas ANOVA and Kruskal-Wallis continued to yield extremely low p-values. This pattern suggests that traditional methods may overstate the evidence for significance in noisy data, while TABS

responds more cautiously. In Simulation-4 and Simulation-5, where multiple groups displayed heterogeneous skewness (e.g., both right- and left-skewed distributions), the TABS test maintained strong discriminative power even as outlier rates increased. Especially in Simulation-5, TABS yielded high test statistics and highly significant p-values (TABS = 814.3, p = 0.000 at 0% outliers; TABS = 528.4, p = 0.000 at 20% outliers), showing its robustness under extreme asymmetry and contamination. These results highlight two key strengths of the TABS test:

1. It is resilient to non-normality and outlier contamination, unlike classical tests whose assumptions may not hold in real-world data.
2. It evaluates both the central tendency and shape of distributions, offering a richer understanding of group differences.

Thus, TABS presents itself not merely as a substitute, but as a complementary and often preferable alternative in complex data environments.

A limitation of this study is the absence of comparison with recently proposed robust alternatives such as quantile ANOVA or WRS2-based effect size frameworks. Integrating these methods in future work will provide a more comprehensive evaluation of TABS's performance relative to state-of-the-art robust approaches.

CONCLUSION

This study proposed a new non-parametric permutation test - the TABS statistic - based on trimean and Bowley's asymmetry measure, aimed at improving group comparison reliability when classical assumptions are violated. The method was tested using both real and simulated datasets, including data contaminated with various proportions of outliers and skewed distributions. Comparisons with ANOVA and Kruskal-Wallis revealed that while traditional methods consistently yielded statistically significant results, the TABS test adapted its behavior depending on the distributional characteristics of the data. Overall, the TABS test provides robust results under skewness, heterogeneity, and small sample conditions, reduces the risk of Type I errors by not overreacting to extreme values and supported by a permutation-based p-value framework, making it free from distributional assumptions.

These findings suggest that the TABS method is a strong alternative to classical multiple comparison tests, especially when working with real-world data that deviate from ideal statistical assumptions. It may be particularly useful in fields like social sciences, biology, or economics, where such irregularities are common.

Future studies may consider a more extensive comparison of the TABS test with recently developed robust statistical frameworks to better contextualize its performance in modern applied settings.

DATA STATEMENT AVAILABILITY

Real-time data supporting the findings of this study (iris, mtcars, PlantGrowth, warpbreaks, chickwts, InsectSprays) are readily available on the R software program and internet sources. Simulation data were generated by the author and are available upon request from the corresponding author.

CONFLICT OF INTERESTS

No potential conflict of interest was reported by the author.

REFERENCES

1. Fisher, R.A. *Advanced Statistical Methods for Research Workers*, Oliver&Boyd, Edinbugh, UK, **1925**.
2. Ross, S. *Probability and statistics for engineers and scientists*, Elsevier, New Delhi, India, **2009**.
3. Şenoğlu, B., and Acıtaş, Ş. *İstatistiksel Deney Tasarımı-Sabit Etkili Modeller*, Nobel Akademik Yayıncılık, Ankara, Türkiye, **2011**.
4. Lindman, H. R. *Analysis of variance in experimental design*, Springer Science & Business Media, **2012**.
5. Montgomery, D. C. *Design and analysis of experiments*, John Wiley & Sons, **2017**.
6. Gravetter, F. J., and Wallnau, L. B. Statistics for the Behavioral Sciences. *Cengage Learning*, **2017**.
7. Field, A. *Discovering Statistics Using IBM SPSS Statistics*, Sage Publications, **2018**.
8. Wilcox, R. R. *Introduction to robust estimation and hypothesis testing*. Academic press, **2011**.
9. Mair, P. and Wilcox, R. R. A unified framework for robust group comparisons using quantile-based effect sizes. The American Statistician, **2022**, 76(3), 230–242.
10. Benjamini, Y. and Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing, *Journal of the Royal statistical society: series B*, **1995**, 57(1), 289-300.
11. Konietschke, F., Bösiger, S., Brunner, E. and Hothorn, L. A. Are multiple contrast tests superior to the ANOVA?, *The International Journal of Biostatistics*, **2013**, 9(1), 63-73.
12. Pallmann, P. and Hothorn, L. A. Analysis of means: a generalized approach using R, *Journal of Applied Statistics*, **2016**, 43(8), 1541-1560.
13. Hothorn, L. A. The two-step approach - a significant ANOVA F-test before Dunnett's comparisons against a control - is not recommended, *Communications in Statistics-Theory and Methods*, **2016**, 45(11), 3332-3343.
14. Rao, C. V. Analysis of means - a review, *Journal of Quality Technology*, **2005**, 37(4), 308-315.
15. Bretz, F., Genz, A. and A. Hothorn, L. On the numerical availability of multiple comparison procedures, *Biometrical Journal*, **2001**, 43(5), 645-656.
16. Gao, X. and Alvo, M. Nonparametric multiple comparison procedures for unbalanced two-way layouts, *Journal of Statistical Planning and Inference*, **2008**, 138(12), 3674-3686.
17. Mahmood, T., Riaz, M., Iqbal, A. and Mulenga, K. An improved statistical approach to compare means, *AIMS Math*, **2023**, 8(2), 4596-4629.

18. Noguchi, K., Abel, R. S., Marmolejo-Ramos, F. and Konietschke, F. Nonparametric multiple comparisons, *Behavior Research Methods*, **2020**, 52, 489-502.
19. Mair, P. and Wilcox, R. Robust statistical methods in R using the WRS2 package, *Behavior research methods,* **2020**, 52, 464-488.
20. Basu, A., Mandal, A., Martín, N. and Pardo, L. A robust wald-type test for testing the equality of two means from log-normal samples, *Methodology and Computing in Applied Probability,* **2019**, 21, 85-107.
21. Hothorn, L. A. and Kluxen, F. M. Robust multiple comparisons against a control group with application in toxicology. *arXiv preprint arXiv:1905.01838*, **2019**.
22. Wickham, H., François, R., Henry, L., Müller, K., Vaughan, D. and PBC. *dplyr: A Grammar of Data Manipulation,* doi: 10.32614/CRAN.package.dplyr, **2025**.
23. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org, **2016**.