Menemui Matematik (Discovering Mathematics) 47(3) (2025) 65-75

Menemui Matematik
(Discovering Mathematics)

journal homepage: https://persama.org.my/dismath/home

Modification on O' Brien Procedure and Anova F-Test Using Winsorized Mean and Variance For A Fixed Effect Two-Factor Factorial Design

Sam Jie Ting[1] and Nazihah Mohamed Ali[2*]
[1,2]*Department of Mathematics and Statistics, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor,*

[1]211219@student.upm.edu.my@upm.edu.my, [2]nazihanma@upm.edu.my
*Corresponding author

Received: 28 July 2025
Accepted: 8 November 2025

ABSTRACT

Variance heterogeneity is a common issue that can distort statistical inference in ANOVA models. However, real-world data often violate the assumption of homogeneity of variance (HOV), which states that the variances across groups are equal. The O' Brien test is a common HOV test that frequently applied to assess variance equality before conducting further analysis. This study proposes a modification of O' Brien test using Winsorized mean and variance as robust measures of central tendency and scale to improve robustness under variance heterogeneity. The performance of the tests is evaluated based on Type I error under various simulation conditions. This study evaluates the performance of original and modified O' Brien test in a 3 × 3 factorial design, followed by a two-factor factorial ANOVA and Duncan's multiple range test (DMRT) under different variance ratios and three distributions (normal, Chi square and lognormal). Simulation results indicate that the modified O' Brien test outperforms the original in handling moderately unequal variances across most conditions. Furthermore, results show that the DMRT under modified ANOVA demonstrates improved robustness in identifying significant mean differences under moderate and extreme variance heterogeneity.

Keywords: Winsorized mean; O' Brien test; Two-factor Factorial ANOVA; Duncan's Multiple Range Test; Type I error

INTRODUCTION

Analysis of Variance (ANOVA), developed by Sir Ranold A. Fisher (1925), is a statistical test to examine the difference in means among multiple groups. It is widely utilized in various experiments, including biology, chemistry, engineering and agriculture. There are three primary assumptions that need to be fulfilled before conducting ANOVA, i.e. normality, independence and homogeneity of variances. Violation of the assumption of HOV may lead to falsely reject a true null hypothesis and thereby increase the probability of Type I error.

Several tests exist for assessing homogeneity of variance, such as Barlett's test, Levene' test and O' Brien test. Barlett's test is powerful but it is sensitive to non-normality assumption, wheras Levene's is more robust to non-normality but can be lose statistical power under certain conditions (Hatchavanich, 2014). The O' Brien test, developed by O' Brien in 1979, is one of the statistical methods used to test variance homogeneity. This test transforms the original scores and applies

ANOVA to the transformed scores. It is comparative with other tests in terms of power and it can be easily applied in different ANOVA designs with equal and unequal sample sizes (Hatchavanich, 2014). However, like other mean-based methods, the O' Brien test remains sensitive to extreme values, which can distort results under non-normal distributions.

Most previous work on HOV tests has focused on one-way ANOVA, while limited attention has been given to the application of O' Brien test and its modifications in factorial designs. To bridge this gap, this study modifies the O' Brien test with robust measures of location and scale, namely Winsorized mean and Winsorized variance. The Winsorization approach maintains the data structure while reducing the effect of extreme values on overall mean when dealing with skewed distributions (Rivest, 1994). Hence, by integrating these modifications, this study aims to address the issue of heterogeneity and improve the robustness of the test in factorial designs.

METHODOLOGY

Original O' Brien Test

The methodology begins with the O' Brien test to assess the equality of variance, followed by the application of two-factor factorial ANOVA and finally post-hoc multiple comparison test.

In this study, O' Brien test is selected to assess the homogeneity of variance among other HOV tests due to its sensitivity, versatility and compatibility to standard ANOVA and minimize both the Type I and Type II error (Abdi, 2007). The hypothesis of O' Brien test can be stated as:

$$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_j^2$$
$$H_1: \sigma_i^2 \neq \sigma_j^2 \text{ for at least one pair } (i, j)$$

The main idea of the O' Brien test is to transform the original scores, y_{ijk} for each sample to the transformed scores, r_{ijk} by using the following formula:

$$r_{ijk} = \frac{(n_{jk} - 1.5)(n_{jk})(y_{ijk} - \bar{y}_{jk})^2 - 0.5s_{jk}^2(n_{jk} - 1)}{(n_{jk} - 1)(n_{jk} - 2)} \tag{1}$$

where r_{ijk} represents the transformed value for an observation y_{ijk}, $\bar{y}_{jk}$ is the group mean, s_{jk}^2 is the group variance and n_{jk} is the number of observations in the group.

The transformation in the O' Brien test converts variance difference into mean difference. In essence, the mean of the transformed values for a particular cell equals to the corresponding cell variances. The O' Brien test statistic is then obtained as the F-value, calculated by applying usual ANOVA procedure for a factorial design to the transformed scores, r_{ijk}.

Modified O' Brien Test with Winsorized Mean and Winsorized Variance

The arithmetic means in original O' Brien test is easily influenced by extreme values and thus is no longer good estimator of location and scale. By applying Winsorization approach, the impact of extreme values is reduced by replacing a fixed percentage of the smallest and the largest observations with the nearest remaining values, paying more attention to data near the center (Wilcox, 2003). Hence, a modification is done by incorporating Winsorized mean and Winsorized variance to the original O' Brien test.

$$r_{ijk} = \frac{(n_{jk} - 1.5)(n_{jk})(y_{ijk} - \bar{y}_{iw.})^2 - 0.5s_{iw.}^2(n_{jk} - 1)}{(n_{jk} - 1)(n_{jk} - 2)} \quad (2)$$

where $\bar{y}_{iw}$ is the 10% Winsorization mean and $s_{iw.}^2$ is the Winsorized variance of the i^{th} subgroup, with the dot indicating the averaging across replicates. According to Hill and Dixon (1982), a 10% Winsorized mean is sufficient for controlling Type I error rates.

Modified Two-factor Factorial ANOVA

Two-factor factorial ANOVA is employed to analyze how two independent variables is influenced by two independent factors, both individually (main effects) and in combination (interaction effect). The statistical model for this design, where factor A has a levels and factor B has b levels, is formally defined as:

$$Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}$$

where μ is the overall mean, while τ_i, β_j and $(\tau\beta)_{ij}$ represent the i^{th} level of factor A, the j^{th} level of factor B and their interaction effect respectively. The term ε_{ijk} denotes the independent and identically distributed (i.i.d.) error component. This framework allows for testing the null hypothesis that there no main effects for each factor and no interaction effect:

$$H_0: \tau_1 = \tau_2 = \cdots = \tau_a$$
$$H_0: \beta_1 = \beta_2 = \cdots = \beta_b$$
$$H_0: (\tau\beta)_{11} = (\tau\beta)_{12} = \cdots = (\tau\beta)_{ab}$$

The modified version of this ANOVA replaces the arithmetic mean with the Winsorized mean. Consequently, the F-statistic for the main and interaction effects are calculated using this robust measure of central tendency. The modified F-statistic for a model incorporating a 10% Winsorized mean are defined as follows:

Row effect (Factor A): $F_A = \frac{MSA_W}{MSE_W}$ (3)

Column effect (Factor B): $F_B = \frac{MSB_W}{MSE_W}$ (4)

Interaction effect ($A \times B$): $F_{AB} = \frac{MSAB_W}{MSE_W}$ (5)

where MSA_W is the Winsorized mean square for factor A, MSB_W is the Winsorized mean square for factor B, $MSAB_W$ is the Winsorized mean square for interaction between factor A and factor B and MSE_W is the Winsorized mean square error.

In the modified two-factor ANOVA, F-statistic are computed based on Winsorized mean. For each effect (A, B or AB), the null hypothesis is rejected at significant level α if its F-statistic exceed the critical value from F-distribution with df_1 and df_2 degrees of freedom. The df_1 represents the degree of freedom for factor A, factor B or their interaction, while df_2 corresponds to the degree of freedom of error. A statistically significant result provides evidence of a main or interaction effect, whereas a non-significant result does not support the presence of such an effect.

Post-Hoc Analysis using Duncan's Multiple Range Test (DMRT)

ANOVA assesses the differences in groups means by examining the F-statistic and its associated p-value. When the null hypothesis is rejected in ANOVA, it indicates that at least one group mean differs significantly from the others, but it does not specify which groups are different. To identify these specific differences, Duncan's Multiple Range Test (DMRT) was employed as a post-hoc procedure. DMRT is favor for its high statistical power and effectiveness in controlling Type II errors (Stoll, 2017). The hypotheses for each pairwise comparison in DMRT are:

H_0: There is no significant different between the means of all possible pairs.
H_1: At least one pair of groups has significant difference in mean.

The formula of DMRT can be defined as follows:

$$DMRT = r_p \times \sqrt{\frac{MSE}{n_k}} \tag{6}$$

where r_p is the least squared studentized range, MSE is the mean squared error and n_k is the sample size of the cell.

The null hypothesis is rejected when the difference in mean, $|\bar{y}_i - \bar{y}_j|$ is greater than the critical value, D_p, where D_p is computed as $D_p = r_P \times S_{\bar{y}_{ij}}$. Hence, it can be concluded that the pairs of means are significantly different from each other.

SIMULATION STUDY

The simulation study is conducted by manipulating the three variables to create different situations for comparing the strengths and weakness of the O' Brien tests, two-factor factorial ANOVA and Duncan's Multiple Range test.

Random samples are generated in R programming according to a 3×3 factorial design, with the simulation conditions summarized in Table 1. For each simulated dataset, the equality of variance is assessed using O' Brien and modified O' Brien test. Subsequently, a two-factor factorial ANOVA is applied on the data. When the interaction effect is found to be significant, Duncan's Multiple Range Test is conducted to examine pairwise group differences. Each simulation is repeated 10000 times and the performance of the statistical test is evaluated based on the Type I error, which serves as a basis for evaluating robustness by referring to Bradley's liberal criterion of robustness. The Type I error is calculated as the proportion of times the p-value falls below the 0.05 significance level out of the 10000 simulations. The three manipulated variables are outlined in Table 1 below:

Table 1: Simulation conditions

Number of replications	$n = 5, N = 45$ $n = 10, N = 90$ $n = 40, N = 360$
Degree of variance heterogeneity	Equal (1:1:1) Moderately unequal (1:1:4) Extremely unequal (1:1:16)
Type of distributions	Normal distribution, $X \sim (\mu, \sigma^2)$ Chi-square distribution, $\chi^2(df)$ Lognormal distribution, $X \sim (\mu, \sigma\ )$

A balanced 3×3 factorial design consists of three level of factor A, three levels of factor B and n independent replications for each 3×3 treatment combinations. The total sample size is given by $N = abn$. In this study, the number of replications per cell is set to $n = 5, 10, 40$, corresponding to total sample size of 45, 90 and 360 respectively.

The degree of variance heterogeneity provides a spectrum ranging from equal to moderately and extremely unequal variances. Variance ratio of 1:1:1, 1:1:4 and 1:1:16 have been widely used by researchers in their simulation studies and serve as benchmark conditions to evaluate the robustness of statistical test under increasing levels of variance heterogeneity (Yonar et al., 2024; Abdullah & Muda, 2022).

For the distribution's selection, normal distribution is an ideal scenario for most statistical tests perform ideally under this distribution and maintain nominal Type I error (Blanca et al., 2023). On the other hand, Chi-square and lognormal distribution are commonly used to represent moderately skewed and heavily skewed distribution, to assess the robustness of statistical tests when the data deviate from normality.

According to Bradley's liberal criterion of robustness (1978), statistical tests are considered as conservative when the Type I error is less than 0.5α, robust when it lies between 0.5α and 1.5α and liberal when it is greater than 1.5α. To provide an easier interpretation and identify the patterns of robustness, these criteria are visually distinguished using distinct color codes. For a significance level of 0.05, the corresponding Type I error ranges, criteria of robustness and color codes are tabulated in Table 2.

Table 2: Type I Error Rates for $\alpha = 0.05$

Bradley's Interval	Type I Error (at $\alpha = 0.05$)	Criteria of Robustness
$\alpha < 0.5\alpha$	$\alpha < 0.025$	Conservative
$0.5\alpha < \alpha < 1.5\alpha$	$0.025 < \alpha < 0.075$	Robust
$\alpha > 1.5\alpha$	$\alpha > 0.075$	Liberal

Robustness is the capability to control the Type I error when the data is not normally distributed and the groups are in fact homogeneous. An overly liberal test may falsely reject the true null hypothesis too often, leading to an increased risk of false positives. Conversely, an overly conservative test may reject the null hypothesis too infrequently, which may reduce the power of the tests (Kamath et al., 2025).

RESULTS AND DISCUSSION

In this study, the main focus is to examine the Type I error in the interaction effect between factor A and factor B, rather than their main effects. In the subsequent interpretations and discussions, our attention is directed solely toward the interaction effect, similar to the approach taken in two-factor factorial ANOVA. For Duncan's Multiple Range Test, the primary interest lies in identifying which means of factor A differ, while holding level 1 of factor B (B1) constant. For instance, the notation "2-1" represents a pairwise comparison between the mean of level 2 and level 1 of Factor A (and vice versa).

Normal Distribution

Table 3: Type I error for Original and Modified Approaches of the O' Brien test and ANOVA Simulated under Normal Distribution

Degree of Variance	Replications	Original		Modified	
		Original O' Brien	Two-factor Factorial ANOVA	O' Brien	Two-factor Factorial ANOVA
1:1:1	$n = 5$	0.0303	0.0514	0.0257	0.0563
	$n = 10$	0.0436	0.0507	0.0405	0.0563
	$n = 40$	0.0472	0.0496	0.0462	0.0580
1:1:4	$n = 5$	0.0532	0.0672	0.0483	0.0709
	$n = 10$	0.0665	0.0654	0.0663	0.0731
	$n = 40$	0.0760	0.0636	0.0741	0.0718
1:1:16	$n = 5$	0.0694	0.0951	0.0664	0.1000
	$n = 10$	0.0796	0.0879	0.0782	0.0962
	$n = 40$	0.0912	0.0801	0.0892	0.0905

Based on the findings in Table 3, under equal (1:1:1) and moderately unequal (1:1:4) variance, the original O' Brien test is robust across various replication sizes, except when the cell size is $n = 40$. Meanwhile, for the modified O' Brien test, it shows robustness for all replication sizes, regardless of small, moderate and large sample sizes. However, under extremely unequal variance ratio (1:1:16), both the original and modified O' Brien test remains robust only for small replication sizes, losing robustness as the replication size increases.

For the two-factor factorial ANOVA, under equal (1:1:1) and moderately unequal (1:1:4) variance, both the original and modified approaches effectively controlling the Type I error within the robust range of 0.025 to 0.075 across varying replication levels. When the variance ratio is extremely unequal (1:1:16), both approaches exhibit a liberal characteristic, regardless of the cell size.

Table 4: Type I Error for Duncan's Multiple Range Test in Classical ANOVA and Modified ANOVA Simulated under Normal Distribution

Degree of Variance	Replications	Classical ANOVA			Modified ANOVA		
		2-1	3-1	3-2	2-1	3-1	3-2

1:1:1	$n = 5$	0.2082	0.2179	0.1926	0.2096	0.1954	0.1954
	$n = 10$	0.2249	0.1952	0.2071	0.2075	0.1941	0.1932
	$n = 40$	0.1854	0.1895	0.1976	0.1793	0.1776	0.1724
1:1:4	$n = 5$	0.0908	0.0908	0.0934	0.0889	0.1001	0.0903
	$n = 10$	0.0934	0.0810	0.0932	0.0807	0.0739	0.0917
	$n = 40$	0.0723	0.0818	0.0881	0.0794	0.0864	0.0780
1:1:16	$n = 5$	0.0578	0.0578	0.0610	0.0610	0.0680	0.0530
	$n = 10$	0.0523	0.0478	0.0569	0.0499	0.0509	0.0509
	$n = 40$	0.0462	0.0437	0.0587	0.0464	0.0486	0.0519

For Duncan's Multiple Range Test (DMRT), when level B1 is fixed, both classical and modified ANOVA fails to maintain control over the Type I error, with a liberal characteristic consistently observed across most of the cell sizes and pairwise comparisons. Under extremely unequal variance, the Type I error is all robust for all pairwise comparisons and cell sizes.

Chi-square Distribution

Table 5: Type I error for Original and Modified Approaches of the O' Brien test and ANOVA Simulated under Chi-square Distribution

Degree of Heterogeneity	Replications	Original		Modified	
		O' Brien	Two-factor Factorial ANOVA	O' Brien	Two-factor Factorial ANOVA
1:1:1	$n = 5$	0.0682	0.0435	0.0517	0.0406
	$n = 10$	0.0662	0.0468	0.0415	0.0403
	$n = 40$	0.0472	0.0465	0.0284	0.0326
1:1:4	$n = 5$	0.0679	0.0627	0.0570	0.0676
	$n = 10$	0.0683	0.0636	0.0556	0.0679
	$n = 40$	0.0697	0.0634	0.0568	0.0753
1:1:16	$n = 5$	0.0752	0.0909	0.0710	0.1044
	$n = 10$	0.0808	0.0867	0.0752	0.1092
	$n = 40$	0.0888	0.0800	0.0828	0.1635

Based on the findings in Table 5, under equal and moderate unequal variance, both the original and modified O' Brien tests are robust across all cell sizes, showing their ability in controlling the Type I error. Under extremely unequal variance, the original approach is robust at $n = 5$, while the modified approach is robust at $n = 5$ and $n = 10$, showing the modified approach is slightly outperforms the original in controlling Type I error.

Similarly, in the two-factor factorial ANOVA, both the original and modified approaches consistently exhibit robust characteristics when the variance is equal and moderately unequal. On the other hand, when facing with extremely unequal variance (1:1:16), both original and modified approach display a liberal characteristic, with all Type I error exceed 0.075 across all cell sizes.

Table 6: Type I Error for Duncan's Multiple Range Test in Classical ANOVA and Modified ANOVA Simulated under Chi-square Distribution

Degree of Variance	Replications	Classical ANOVA			Modified ANOVA		
		2-1	3-1	3-2	2-1	3-1	3-2

1:1:1	$n = 5$	0.2207	0.2092	0.2067	0.2241	0.2167	0.2044
	$n = 10$	0.1816	0.2030	0.1966	0.1935	0.2208	0.1960
	$n = 40$	0.2000	0.1677	0.1763	0.2331	0.1564	0.1871
1:1:4	$n = 5$	0.0654	0.0845	0.0877	0.0636	0.0814	0.0873
	$n = 10$	0.0723	0.0629	0.0896	0.0751	0.0663	0.0854
	$n = 40$	0.0962	0.0678	0.0931	0.0753	0.0714	0.0805
1:1:16	$n = 5$	0.0431	0.0422	0.0597	0.0431	0.0450	0.0565
	$n = 10$	0.0565	0.0392	0.0565	0.0513	0.0394	0.0513
	$n = 40$	0.0663	0.0350	0.0600	0.0532	0.0355	0.0508

The performance of DMRT under classical ANOVA and modified ANOVA displays a liberal characteristic for all pairwise comparisons when factor B is fixed at B1. Under moderately unequal variance, the DMRT in modified ANOVA demonstrates slightly better control of Type I error compared to classical ANOVA, particularly for the moderate ($n = 10$) and large cell size ($n = 40$). For the extremely unequal variance, the DMRT in classical and modified ANOVA showing its ability to control the Type I error within the range between 0.025 and 0.075, indicating that all the Type I error is robust for all three pairwise comparison, regardless of the cell size.

Lognormal Distribution

Table 7: Type I error for Original and Modified Approaches of the O' Brien test and ANOVA Simulated under Lognormal Distribution

Degree of Variance	Replications	Original		Modified	
		Original O' Brien	Two-factor Factorial ANOVA	Modified O' Brien	Two-factor Factorial ANOVA
1:1:1	$n = 5$	0.0662	0.0349	0.0481	0.1235
	$n = 10$	0.0498	0.0386	0.0333	0.2087
	$n = 40$	0.0351	0.0461	0.0222	0.3040
1:1:4	$n = 5$	0.0742	0.0571	0.0626	0.1613
	$n = 10$	0.0729	0.0611	0.0561	0.2062
	$n = 40$	0.0541	0.0591	0.0384	0.2423

Table 7 (Continued): Type I error for Original and Modified Approaches of the O' Brien test and ANOVA Simulated under Lognormal Distribution

Degree of Variance	Replications	Original		Modified	
		Original O' Brien	Two-factor Factorial ANOVA	Modified O' Brien	Two-factor Factorial ANOVA
1:1:16	$n = 5$	0.0781	0.0889	0.0724	0.2084
	$n = 10$	0.0865	0.0872	0.0748	0.2155
	$n = 40$	0.0826	0.0777	0.0714	0.2153

Based on the findings in Table 7, the original O' Brien test is robust across all replication sizes under equal and moderately unequal variance. However, its Type I error is liberal under extremely unequal variance condition. In contrast, the modified O' Brien test shows a better performance in controlling the Type error across the levels of variance heterogeneity.

Consequently, in the two-factor factorial ANOVA, the original approach demonstrates superior control over Type I error, where all Type I errors in classical ANOVA are robust across different cell sizes, except under extremely unequal variance. In contrast, the modified approach fails to control the Type I error in modified ANOVA, as all values exceed 0.075, indicating a liberal result across all cell sizes.

Table 8: Type I Error for Duncan's Multiple Range Test in Classical ANOVA and Modified ANOVA Simulated under Chi-square Distribution

Degree of Variance	Replications	Classical ANOVA			Modified ANOVA		
		2-1	3-1	3-2	2-1	3-1	3-2
1:1:1	$n = 5$	0.1805	0.1862	0.1834	0.1279	0.1198	0.1142
	$n = 10$	0.1891	0.2021	0.1831	0.0934	0.0939	0.0886
	$n = 40$	0.1844	0.1887	0.1887	0.0796	0.0849	0.0862
1:1:4	$n = 5$	0.0490	0.0595	0.0525	0.0502	0.0477	0.0453
	$n = 10$	0.0687	0.0589	0.0638	0.0589	0.0538	0.0519
	$n = 40$	0.0764	0.0778	0.0795	0.0590	0.0574	0.0561
1:1:16	$n = 5$	0.0337	0.0304	0.0349	0.0326	0.0312	0.0288
	$n = 10$	0.0424	0.0378	0.0390	0.0371	0.0325	0.0312
	$n = 40$	0.0425	0.0476	0.0412	0.0409	0.0432	0.0432

Under equal variance, the DMRT in the classical and modified ANOVA generally exhibits a liberal Type I for all cell sizes. On the other hand, under moderately unequal variance, the DMRT in classical ANOVA displays a robust characteristic for the three pairwise comparisons across different cell sizes, except for the large cell size ($n = 40$). In contrast, the DMRT in modified ANOVA consistently maintains robustness across all comparisons and cell sizes. As for the DMRT in classical and modified ANOVA, all Type I errors are robust for all three comparisons of means across different cell sizes.

CONCLUSION

To conclude, this study provides insights into the comparative efficacy of the original and modified O' Brien test, followed by the two-factor factorial ANOVA and DMRT across different distributions and degree of heterogeneity.

In the normal distribution, the modified O'Brien test controls Type I error well and outperforms the original under moderate variance inequality (1:1:4). Both tests are robust for interaction effects in two-factor factorial ANOVA under equal (1:1:1) and moderate (1:1:4) variance ratios but fail under extremely unequal variance (1:1:16). DMRT remains robust for both classical and modified ANOVA when factor B is fixed at level B1 under moderately unequal variance (1:1:4).

In the Chi-square distribution, both the original and modified approach in O' Brien and two-factor factorial ANOVA, are robust under equal (1:1:1) and moderately unequal (1:1:4) variance ratio. The DMRT under modified ANOVA outperforms classical ANOVA for moderate variance, while under extreme variance (1:1:16), DMRT remains robust for both classical and modified ANOVA.

For the lognormal distribution, the original O'Brien test is robust only under equal variance, whereas the modified O' Brien test maintains robustness under both moderate (1:1:4) and extreme (1:1:16) variance ratios. For the two-factor factorial ANOVA, classical ANOVA is robust under equal (1:1:1) and moderate (1:1:4) variance ratios but liberal under extremely unequal variance (1:1:16). DMRT in both classical and modified ANOVA tends to remain robust under moderate (1:1:4) and extreme (1:1:16) variance ratios.

Overall, the simulation results emphasizes that the modified O' Brien test offers an improvement over the original, as it exhibited exceptional robustness, particularly under moderately and extremely unequal variance ratios. It is important to note that the results are based on simulation studies and not on real data. Future work of this study can be further extended by incorporating other robust measures such as trimmed mean into the O' Brien formula, to compare its performance with the current study.

REFERENCES

Abdi, H. (2007). O'Brien test for homogeneity of variance. *Encyclopedia of measurement and statistics,* **2**: 701-704.

Blanca-Mena, M. J., Arnau, J., García-Castro, F. J., Alarcón-Postigo, R., & Bono Cabré, R. (2022). Non-normal Data in Repeated Measures ANOVA: Impact on Type I Error and Power.

Bradley, J. V. (1978). Robustness?. *British Journal of Mathematical and Statistical Psychology,* **31(2)**: 144-152.

Dixon, W. J., & Yuen, K. K. (1974). Trimming and winsorization: A review. *Statistische Hefte,* **15(2)**: 157-170.

Duncan, D. B. (1955). Multiple range and multiple F tests. *biometrics*, *11*(1), 1-42.

Hatchavanich, D. (2014). A comparison of type I error and power of Bartlett's test, Levene's test and O'Brien's test for homogeneity of variance tests. *Southeast Asian Journal of Sciences*, **3(2)**: 181-194.

Kamath, A., Poojari, S., & Varsha, K. (2025). Assessing the robustness of normality tests under varying skewness and kurtosis: a practical checklist for public health researchers. *BMC Medical Research Methodology*, **25(1)**: 206.

O'Brien, R. G. (1978). Robust techniques for testing heterogeneity of variance effects in factorial designs. *Psychometrika*, **43(3)**: 327-342.

O'Brien, R. G. (1981). A simple test for variance effects in experimental designs. *Psychological Bulletin*, **89(3)**: 570.

Ujian, K., Abdullah, N. F., & MUDA, N. (2022). An overview of homogeneity of variance tests on various conditions based on type 1 error rate and power of a test. *Journal of Quality Measurement and Analysis JQMA*, **18(3)**: 111-130.

Wilcox, R. R. (2003). 3-summarizing data. *Applying Contemporary Statistical Techniques*, 55-91.

Wilcox, R. R. (2011). *Introduction to robust estimation and hypothesis testing*. Academic press.

Yi, Z., Chen, Y. H., Yin, Y., Cheng, K., Wang, Y., Nguyen, D., ... & Kim, E. (2022). Brief research report: A comparison of robust tests for homogeneity of variance in factorial ANOVA. *The Journal of Experimental Education*, **90(2)**: 505-520.

Yonar, A., Yonar, H., Demirsöz, M., & Tekindal, M. A. (2024). A COMPARATIVE ANALYSIS FOR HOMOGENEITY OF VARIANCE TESTS. *Journal of Science and Arts*, **24(2)**: 305-328..

Yuen, K. K. (1971). A note on Winsorized t. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **20(3)**: 297-304.