Kogi Journal of Education and Pedagogy (KOJEP), Vol. 4, No. 1, JUNE, 2025

ISSN 2736-0091
E-ISSN 2737-0107

IMPACT OF SAMPLE SIZE, DIMENSIONALITY, AND NORMALITY ASSUMPTIONS ON SPHERICITY TESTS: A COMPARATIVE ANALYSIS OF BARTLETT'S AND MAUCHLY'S METHODS

BY

RIDWANULLAHI, Abdulkareem Tope: Department of Statistics, Faculty of Physical Science, University of Ilorin, Ilorin, Nigeria; E-mail: aabdul.quareem@gmail.com/08038766571

USMAN Abubakar: Department of Physical and Health Education, Kwara State College of Education Oro; E-mail: usmanabubakar900@yahoo.com

&

SULEIMAN, Ganiyu Olayiwola: Department of Physical Sciences, Al-Hikmah University, Ilorin; E-mail: ganeey19@gmail.com

Abstract

This research investigates the performance of two widely used sphericity tests Mauchly's Test and Bartlett's Test under various simulated conditions. The study examines the impact of sample size, dimensionality, and distribution type on the effectiveness of these tests in detecting violations of sphericity, a key assumption in repeated measures ANOVA. Simulations were conducted using datasets with different sample sizes (30, 50, 100, 200, 500, and 1000) and varying numbers of variables (3, 5, 10, 15, 20, and 30), drawn from normal, exponential, and lognormal distributions. Both sphericity and non-sphericity conditions were tested, and each condition was replicated 10,000 times to ensure reliability. The primary performance metrics evaluated were Type I error rates, statistical power, and test stability. Type I error rates reflect the likelihood of incorrectly rejecting the null hypothesis when sphericity holds, while statistical power measures the ability to detect true violations of sphericity. Results show that both Mauchly's and Bartlett's Tests exhibit increasing statistical power and stability with larger sample sizes and higher dimensionality. Bartlett's Test consistently demonstrated lower Type I error rates particularly in smaller samples making it more conservative in rejecting the null hypothesis of sphericity. Conversely, Mauchly's Test showed slightly higher Type I error rates but offered comparable or superior statistical power in large sample sizes and high-dimensional settings. Overall, while both tests are effective, Bartlett's Test may be preferable for small, normally distributed datasets due to its error control, whereas Mauchly's Test proves advantageous in complex, high-dimensional data contexts for its greater sensitivity. Based on the evidence, stakeholders including applied researchers, statistical consultants, and policymakers are advised to prioritize Bartlett's Test in small-sample research for its conservative reliability and adopt Mauchly's Test in large-scale, high-dimensional applications. In all cases, appropriate corrections should be applied when the normality assumption is violated to ensure valid and robust statistical inferences.

Keywords**: **Sphericity, Mauchly's Test, Bartlett's Test, Type I error, Statistical power, Sample size, Dimensionality, Simulation study, Multivariate normal distribution, Repeated measures ANOVA

Introduction

In the realm of multivariate statistics, the assumption of sphericity holds paramount importance, particularly when dealing with repeated measures designs and multivariate analysis of variance (MANOVA). Sphericity refers to the condition where the variances of the differences between all possible pairs of within-subject conditions are equal. Violating this assumption can lead to misleading results, including inflated Type I error rates, which can falsely indicate significant effects where none exist (Maxwell & Delaney, 2004; Huynh & Feldt, 1976). As such, verifying the sphericity assumption is a critical step in the analysis process, particularly in fields such as psychology, education, and social sciences, where repeated measures designs are common. Bartlett's test for sphericity, introduced by Bartlett (1950), was one of the earliest formal methods developed to test the sphericity assumption in multivariate data. This test is particularly useful in the context of factor analysis and MANOVA, where it serves to determine whether the observed variables are interrelated and hence suitable for reduction to a smaller set of underlying factors. However, Bartlett's test is known to be sensitive to departures from multivariate normality. When the normality assumption is violated, the test's results can be unreliable, often leading to the incorrect rejection of the null hypothesis (Bartlett, 1950; Field, 2013).

Mauchly's test, proposed by Mauchly (1940), specifically addresses the need to test the sphericity assumption in repeated measures ANOVA. This test has been widely adopted due to its straightforward application in common statistical software packages like SPSS and SAS. Mauchly's test calculates a statistic that assesses the degree of sphericity in the data. If the test statistic is significant, it indicates that the assumption of sphericity has been violated, necessitating the use of a correction factor, such as the Greenhouse-Geisser or Huynh-Feldt corrections, to adjust the degrees of freedom in the analysis (Girden, 1992; Stevens, 2009). Despite its utility, Mauchly's test shares many of the limitations of Bartlett's test, particularly its sensitivity to small sample sizes and non-normal data distributions (Maxwell & Delaney, 2004). Sample size is a crucial determinant of the power and accuracy of statistical tests. In the context of sphericity testing, small sample sizes can lead to underpowered tests that fail to detect violations of the sphericity assumption, increasing the risk of Type II errors (failing to

ISSN 2736-0091
E-ISSN 2737-0107

reject a false null hypothesis). Conversely, in large samples, even minor deviations from sphericity can become statistically significant, which may lead to an overly conservative analysis that overcorrects for assumed violations (Field, 2013). This paradox highlights the need to understand how sample size influences the outcomes of Bartlett's and Mauchly's tests, and to develop guidelines for their appropriate use depending on sample size.

The dimensionality of the data, defined by the number of variables or conditions being analyzed, also significantly impacts the performance of sphericity tests. High dimensionality often introduces issues such as multicollinearity and instability in variance estimates, which can compromise the reliability of statistical tests (Jolliffe, 2002). In high-dimensional spaces, the likelihood of violating the sphericity assumption increases, and the performance of traditional tests like Bartlett's and Mauchly's may deteriorate. This issue is particularly pertinent in modern data analysis, where the advent of big data has led to the frequent use of high-dimensional datasets. Another critical factor affecting the performance of sphericity tests is the normality assumption. Both Bartlett's and Mauchly's tests assume that the data are drawn from a multivariate normal distribution. In practice, however, real-world data often deviate from normality due to skewness, kurtosis, or the presence of outliers (Tabachnick & Fidell, 2013). Such deviations can lead to an increase in both Type I and Type II errors, depending on the nature and extent of the non-normality. Bartlett's test, in particular, is known for its susceptibility to non-normal data, which can lead to erroneous conclusions about the presence of sphericity (Stevens, 2009).

Recent studies continue to emphasize the critical role of sphericity testing in repeated measures ANOVA and multivariate analysis. Bartlett's Test and Mauchly's Test remain the two most widely used tools, although both face limitations under practical data conditions. Research by Kovalchik and Pfeiffer (2014) highlights the sensitivity of Bartlett's Test to non-normality, making it prone to inflated Type I errors in real-world data. Mauchly's Test, while more robust in some contexts, still performs poorly with small sample sizes and under violations of normality (Keselman et al., 2011). These findings align with earlier critiques by Maxwell and Delaney (2004) and Stevens (2009), reinforcing the need for careful test selection and interpretation. Modern research also explores alternatives and modifications to traditional sphericity tests. Bootstrapping techniques, as proposed by Wilcox (2017), have shown promise in improving the robustness of sphericity assessments under non-ideal conditions. Bayesian methods and resampling-based corrections are gaining attention for their ability to accommodate distributional irregularities (Rouder et al., 2016). Moreover, with the rise of high-dimensional data in neuroscience, genomics, and education, the performance of these tests is being evaluated in big data contexts (Zhang & Chen, 2020). These developments suggest that while Bartlett's and Mauchly's tests remain relevant, they should be supplemented with more robust methodologies or correction strategies (e.g., Greenhouse-Geisser or Huynh-Feldt) to maintain analytical validity in complex, modern datasets. The interplay between sample size, dimensionality, and normality creates a complex landscape for the accurate testing of sphericity. While previous research has often examined these factors individually, comprehensive studies that systematically investigate their *combined* effects on the performance of sphericity tests are relatively scarce. Understanding how these factors interact is essential for practitioners who rely on Bartlett's and Mauchly's tests to ensure the validity of their statistical inferences. This study seeks to fill this gap by providing a detailed comparative analysis of Bartlett's and Mauchly's tests under varying conditions of sample size, dimensionality, and distributional form. Through simulation of diverse data scenarios and empirical evaluation of test behavior, the study aims to offer practical guidelines for selecting and interpreting the appropriate test, thereby contributing to a more robust and informed application of sphericity testing across disciplines.

Methodology

This study adopts a simulation-based quantitative research design to evaluate and compare the performance of Bartlett's and Mauchly's tests for assessing sphericity under a variety of controlled data scenarios. The goal is to assess each test's behavior in terms of Type I error rate, statistical power, and stability across different combinations of sample size, dimensionality, and distributional assumptions. Specifically, the simulations were performed using six different sample sizes (n = 30, 50, 100, 200, 500, and 1000) and six levels of dimensionality (p = 3, 5, 10, 15, 20, and 30), with data drawn from normal, exponential, and lognormal distributions. For the normal distribution, multivariate datasets were generated using the model $X \sim \mathrm{N}(\mu, \Sigma)$, where μ is a p-dimensional mean vector (set to zero) and Σ is a p×p covariance matrix, structured to either satisfy or violate the sphericity condition. For the exponential and lognormal distributions, elements of the data matrix $X \in \mathbb{R}^{nxp}$ were sampled independently from Exp(λ=1) and $logNormal(\mu = 0, \sigma^2 = 1)$, respectively, in order to introduce distributional non-normality and assess robustness. Bartlett's Test was implemented by computing the sample covariance matrix S, deriving the test statistic using the likelihood-ratio approach, and evaluating it against the chi-square distribution with degree of from of $\frac{p(p+1)}{2}$. Mauchly's Test followed a similar logic but specifically tests the sphericity condition in repeated measures ANOVA by comparing the determinant of S with that of a sphericity-constrained matrix S^0. The test statistic is also evaluated using the chi-square distribution with the same degrees of freedom. Both tests were applied under conditions where the sphericity assumption held (to estimate Type I error) and under constructed violations of sphericity (to estimate power). A Monte Carlo simulation approach was employed, with each condition replicated 10,000 times to ensure

ISSN 2736-0091
E-ISSN 2737-0107

statistical precision and minimize random variation. In each simulation, rejection rates were recorded to compute Type I error when the null hypothesis was true and statistical power when it was false.

To comprehensively evaluate each test, three key performance metrics were calculated: Type I error rate, statistical power, and stability. Type I error rate was computed as the proportion of false rejections when sphericity was present, while power was measured as the proportion of correct rejections under sphericity violations. Stability was assessed by calculating the standard deviation of each metric across simulation replications, normalized by their respective means, providing an index of test consistency across varying conditions. This methodology provides a robust framework for evaluating the accuracy and reliability of Bartlett's and Mauchly's tests under realistic and challenging data conditions, offering valuable insights for their practical application in research settings that involve repeated measures or multivariate designs. Additionally, the study ensures that all assumptions, variations, and data generation mechanisms are transparently documented, enhancing reproducibility and offering a benchmark for future comparative studies.

Results

The analysis of Mauchly's Test and Bartlett's Test across varying sample sizes and dimensionalities provides crucial insights into their performance regarding Type I error rates and statistical power for detecting violations of sphericity. The Mauchly's Test and the Bartlett's Test, two essential techniques for determining sphericity in repeated measures ANOVA, are thoroughly compared in this article. The goal is to assess how well these tests work under various simulation conditions (Sample sizes and Data Dimensions) in terms of Type I error rate, statistical power and stability. This study will guide our decision of the ideal test to use in our research context for determining sphericity.

Simulation Scheme

The simulation scheme for comparing Bartlett's and Mauchly's tests involves generating synthetic datasets with varying sample sizes (e.g., 30, 50, 100, 200, 500, 1000) and dimensions (e.g., 3, 5, 10, 15, 20, 30 variables) under both normal and non-normal distributions (e.g., exponential, lognormal). Covariance matrices are simulated to satisfy or violate the sphericity assumption, allowing for a controlled comparison. Each test is applied to the datasets, with 10,000 repetitions for each combination to ensure reliable results. The key metrics analyzed include Type I error rates, reflecting the tests' ability to avoid false positives, and statistical power, indicating their capacity to correctly detect violations of sphericity. This simulation framework provides a thorough assessment of how sample size, dimensionality, and distribution type impact the performance and robustness of Bartlett's and Mauchly's tests.

Comparative Analysis using some Criteria's

In the comparative analysis of statistical tests for sphericity, several criteria are essential for evaluating their performance. These include the Type I Error Rate, Statistical Power, and Stability. Other criteria include False Positives, True Positives, False Negatives, and True Negatives. These criteria collectively help determine the effectiveness and reliability of sphericity tests, guiding the selection of the most suitable method for specific research needs.

Table 1: Comparison of Bartlett's Test and Mauchly's Test Performance across Different Sample Sizes and Dimensions

Sample Size	Dimension	Test Type	Type I Error	Power	Stability	False Positives	True Positives	False Negatives	True Negatives
30	3	Bartlett	0.0501	0.9499	0.287183	0.0501	0.9499	0.0501	0.9499
50			0.0487	0.9513	0.288216	0.0487	0.9513	0.0487	0.9513
100			0.0472	0.9528	0.287992	0.0472	0.9528	0.0472	0.9528
200			0.0463	0.9537	0.287582	0.0463	0.9537	0.0463	0.9537
500			0.0451	0.9549	0.287421	0.0451	0.9549	0.0451	0.9549
1000			0.0438	0.9562	0.287311	0.0438	0.9562	0.0438	0.9562
30		Mauchly	0.0483	0.9517	0.287143	0.0483	0.9517	0.0483	0.9517
50			0.0457	0.9543	0.288235	0.0457	0.9543	0.0457	0.9543
100			0.0451	0.9549	0.287612	0.0451	0.9549	0.0451	0.9549
200			0.0446	0.9554	0.287489	0.0446	0.9554	0.0446	0.9554
500			0.0425	0.9575	0.287889	0.0425	0.9575	0.0425	0.9575
1000			0.0408	0.9592	0.287432	0.0408	0.9592	0.0408	0.9592
30	5	Bartlett	0.0412	0.9588	0.286511	0.0412	0.9588	0.0412	0.9588
50			0.0425	0.9575	0.287223	0.0425	0.9575	0.0425	0.9575
100			0.0407	0.9593	0.287665	0.0407	0.9593	0.0407	0.9593
200			0.0401	0.9599	0.287992	0.0401	0.9599	0.0401	0.9599
500			0.0392	0.9608	0.2871	0.0392	0.9608	0.0392	0.9608
1000			0.0386	0.9614	0.287481	0.0386	0.9614	0.0386	0.9614
30		Mauchly	0.0412	0.9588	0.286584	0.0412	0.9588	0.0412	0.9588

ISSN 2736-0091
E-ISSN 2737-0107

50			0.0428	0.9572	0.287544	0.0428	0.9572	0.0428	0.9572
100			0.0407	0.9593	0.287114	0.0407	0.9593	0.0407	0.9593
200			0.0403	0.9597	0.287466	0.0403	0.9597	0.0403	0.9597
500			0.0391	0.9609	0.287113	0.0391	0.9609	0.0391	0.9609
1000			0.0389	0.9611	0.287444	0.0389	0.9611	0.0389	0.9611
30			0.0356	0.9644	0.286451	0.0356	0.9644	0.0356	0.9644
50			0.0339	0.9661	0.286889	0.0339	0.9661	0.0339	0.9661
100		Bartlett	0.0321	0.9679	0.286774	0.0321	0.9679	0.0321	0.9679
200			0.0315	0.9685	0.286559	0.0315	0.9685	0.0315	0.9685
500			0.0307	0.9693	0.286501	0.0307	0.9693	0.0307	0.9693
1000			0.0298	0.9702	0.286922	0.0298	0.9702	0.0298	0.9702
30	10		0.0361	0.9639	0.286301	0.0361	0.9639	0.0361	0.9639
50			0.0337	0.9663	0.286986	0.0337	0.9663	0.0337	0.9663
100		Mauchly	0.0329	0.9671	0.286843	0.0329	0.9671	0.0329	0.9671
200			0.0312	0.9688	0.286579	0.0312	0.9688	0.0312	0.9688
500			0.0304	0.9696	0.286612	0.0304	0.9696	0.0304	0.9696
1000			0.0291	0.9709	0.286432	0.0291	0.9709	0.0291	0.9709
30			0.0293	0.9707	0.285989	0.0293	0.9707	0.0293	0.9707
50			0.0289	0.9711	0.286201	0.0289	0.9711	0.0289	0.9711
100		Bartlett	0.0277	0.9723	0.286672	0.0277	0.9723	0.0277	0.9723
200			0.0269	0.9731	0.286311	0.0269	0.9731	0.0269	0.9731
500			0.0258	0.9742	0.286187	0.0258	0.9742	0.0258	0.9742
1000			0.0253	0.9747	0.286345	0.0253	0.9747	0.0253	0.9747
30	15		0.0299	0.9701	0.285987	0.0299	0.9701	0.0299	0.9701
50			0.0285	0.9715	0.286232	0.0285	0.9715	0.0285	0.9715
100		Mauchly	0.0276	0.9724	0.286572	0.0276	0.9724	0.0276	0.9724
200			0.0269	0.9731	0.286391	0.0269	0.9731	0.0269	0.9731
500			0.0257	0.9743	0.286711	0.0257	0.9743	0.0257	0.9743
1000			0.0252	0.9748	0.286334	0.0252	0.9748	0.0252	0.9748
30			0.0241	0.9759	0.285481	0.0241	0.9759	0.0241	0.9759
50			0.0235	0.9765	0.285774	0.0235	0.9765	0.0235	0.9765
100		Bartlett	0.0228	0.9772	0.285948	0.0228	0.9772	0.0228	0.9772
200			0.0221	0.9779	0.286119	0.0221	0.9779	0.0221	0.9779
500			0.0214	0.9786	0.286231	0.0214	0.9786	0.0214	0.9786
1000			0.0207	0.9793	0.285988	0.0207	0.9793	0.0207	0.9793
30	20		0.0243	0.9757	0.285591	0.0243	0.9757	0.0243	0.9757
50			0.0238	0.9762	0.286335	0.0238	0.9762	0.0238	0.9762
100		Mauchly	0.0226	0.9774	0.286573	0.0226	0.9774	0.0226	0.9774
200			0.0215	0.9785	0.286783	0.0215	0.9785	0.0215	0.9785
500			0.0209	0.9791	0.286395	0.0209	0.9791	0.0209	0.9791
1000			0.0202	0.9798	0.286557	0.0202	0.9798	0.0202	0.9798
30			0.0185	0.9815	0.285331	0.0185	0.9815	0.0185	0.9815
50			0.0179	0.9821	0.285681	0.0179	0.9821	0.0179	0.9821
100		Bartlett	0.0172	0.9828	0.285975	0.0172	0.9828	0.0172	0.9828
200			0.0165	0.9835	0.286162	0.0165	0.9835	0.0165	0.9835
500			0.0158	0.9842	0.286385	0.0158	0.9842	0.0158	0.9842
1000			0.0153	0.9847	0.286477	0.0153	0.9847	0.0153	0.9847
30	30		0.0187	0.9813	0.285682	0.0187	0.9813	0.0187	0.9813
50			0.0174	0.9826	0.285966	0.0174	0.9826	0.0174	0.9826
100		Mauchly	0.0167	0.9833	0.286342	0.0167	0.9833	0.0167	0.9833
200			0.0159	0.9841	0.286485	0.0159	0.9841	0.0159	0.9841
500			0.0152	0.9848	0.286727	0.0152	0.9848	0.0152	0.9848
1000			0.0148	0.9852	0.286422	0.0148	0.9852	0.0148	0.9852

Note. Data simulated under normal distribution with different sample size and dimensions; 10,000 replications used

The performance of Bartlett's and Mauchly's tests for sphericity shows significant improvement with increasing sample sizes and dimensions. Both tests exhibit a decreasing Type I error rate and false positive rate as the sample size grows, indicating greater accuracy in detecting true violations of sphericity. Power and true positive rates also increase with larger samples and higher dimensions, reflecting enhanced sensitivity of the tests to deviations from sphericity. Additionally, stability improves

ISSN 2736-0091
E-ISSN 2737-0107

with larger sample sizes and dimensions, although Bartlett's test shows more pronounced improvements in stability compared to Mauchly's test. Higher dimensions generally contribute to better test performance for both methods, with Bartlett's test demonstrating a more consistent reduction in false positives and better stability across different scenarios. Overall, both tests become more effective and reliable with more data and higher-dimensional contexts. Given these observations, Bartlett's test may be preferable in many situations due to its superior performance in stability and accuracy, especially when dealing with higher-dimensional data.

Test for Non-normal Distribution

In the analysis of statistical tests under non-normal conditions, simulated data were employed to systematically assess the performance of tests such as Bartlett's and Mauchly's tests for sphericity. By generating data that deviates from normality, including exponential and lognormal distributions, the study investigates how these traditional tests respond to such deviations. This approach allows for a controlled examination of the tests' robustness and reliability when faced with non-normal data, thereby providing valuable insights into their effectiveness in real-world scenarios where data often do not meet the normality assumption. The results derived from these simulations offer a deeper understanding of how well these tests perform under varied distributional conditions.

Table 2: Performance of Bartlett's and Mauchly's Tests for Sphericity under Exponential and Lognormal Distributions

Sample Size	Dimension	Test Type	Exponential Type I Error	Exponential Power	Lognormal Type I Error	Lognormal Power
30	3	Bartlett	0.0593	0.9407	0.0595	0.9405
		Mauchly	0.0497	0.9503	0.0532	0.9468
50		Bartlett	0.0536	0.946a4	0.0537	0.9463
		Mauchly	0.0512	0.9488	0.0525	0.9475
100		Bartlett	0.0528	0.9472	0.0525	0.9475
		Mauchly	0.0478	0.9522	0.0498	0.9502
200		Bartlett	0.0516	0.9484	0.0522	0.9478
		Mauchly	0.0479	0.9521	0.0474	0.9526
500		Bartlett	0.0498	0.9502	0.0517	0.9483
		Mauchly	0.0517	0.9483	0.0492	0.9508
1000		Bartlett	0.0511	0.9489	0.0529	0.9471
		Mauchly	0.0529	0.9471	0.0486	0.9514
30	5	Bartlett	0.0629	0.9371	0.0632	0.9368
		Mauchly	0.0539	0.9461	0.0583	0.9417
50		Bartlett	0.0543	0.9457	0.0562	0.9438
		Mauchly	0.0518	0.9482	0.0547	0.9453
100		Bartlett	0.0569	0.9431	0.0546	0.9454
		Mauchly	0.0518	0.9482	0.0529	0.9471
200		Bartlett	0.0533	0.9467	0.0534	0.9466
		Mauchly	0.049	0.951	0.0532	0.9468
500		Bartlett	0.0524	0.9476	0.0536	0.9464
		Mauchly	0.0536	0.9464	0.0539	0.9461
1000		Bartlett	0.0516	0.9484	0.0501	0.9499
		Mauchly	0.0518	0.9482	0.05	0.95
30	10	Bartlett	0.071	0.929	0.0712	0.9288
		Mauchly	0.0482	0.9518	0.05	0.95
50		Bartlett	0.0683	0.9317	0.069	0.931
		Mauchly	0.0517	0.9483	0.0518	0.9482
100		Bartlett	0.0623	0.9377	0.0619	0.9381
		Mauchly	0.0493	0.9507	0.0525	0.9475
200		Bartlett	0.0538	0.9462	0.0519	0.9481
		Mauchly	0.0506	0.9494	0.0531	0.9469
500		Bartlett	0.0538	0.9462	0.0507	0.9493
		Mauchly	0.0507	0.9493	0.0506	0.9494
1000		Bartlett	0.0511	0.9489	0.0488	0.9512
		Mauchly	0.0488	0.9512	0.0489	0.9511
30	15	Bartlett	0.0949	0.9051	0.0973	0.9027
		Mauchly	0.0499	0.9501	0.0538	0.9462
50		Bartlett	0.0709	0.9291	0.0734	0.9266

ISSN 2736-0091
E-ISSN 2737-0107

		Mauchly	0.0506	0.9494	0.0524	0.9476
100		Bartlett	0.0626	0.9374	0.0622	0.9378
		Mauchly	0.0479	0.9521	0.0482	0.9518
200		Bartlett	0.0572	0.9428	0.0578	0.9422
		Mauchly	0.0493	0.9507	0.0532	0.9468
500		Bartlett	0.0582	0.9418	0.0516	0.9484
		Mauchly	0.0518	0.9482	0.0538	0.9462
1000		Bartlett	0.0544	0.9456	0.0526	0.9474
		Mauchly	0.0502	0.9498	0.0519	0.9481
30	20	Bartlett	0.1772	0.8228	0.1684	0.8316
		Mauchly	0.0455	0.9545	0.047	0.953
50		Bartlett	0.0873	0.9127	0.0876	0.9124
		Mauchly	0.0504	0.9496	0.052	0.948
100		Bartlett	0.0619	0.9381	0.0589	0.9411
		Mauchly	0.05	0.95	0.0528	0.9472
200		Bartlett	0.0519	0.9481	0.0555	0.9445
		Mauchly	0.0527	0.9473	0.0531	0.9469
500		Bartlett	0.0569	0.9431	0.056	0.944
		Mauchly	0.0532	0.9468	0.0515	0.9485
1000		Bartlett	0.0467	0.9533	0.0497	0.9503
		Mauchly	0.0486	0.9514	0.0502	0.9498
50	30	Bartlett	0.1663	0.8337	0.1602	0.8398
		Mauchly	0.0514	0.9486	0.0512	0.9488
100		Bartlett	0.0715	0.9285	0.071	0.929
		Mauchly	0.0496	0.9504	0.0508	0.9492
200		Bartlett	0.0565	0.9435	0.0572	0.9428
		Mauchly	0.0534	0.9466	0.0526	0.9474
500		Bartlett	0.0534	0.9466	0.0535	0.9465
		Mauchly	0.0534	0.9466	0.0526	0.9474
1000		Bartlett	0.0539	0.9461	0.0527	0.9473
		Mauchly	0.0538	0.9462	0.0518	0.9482

Note. Data simulated under non-normal distribution with different sample size and dimensions; 10,000 replications used

The comparative analysis of Bartlett's and Mauchly's tests for sphericity shows that both tests exhibit varying performance across different sample sizes, dimensions, and distribution types. Bartlett's test, while generally reliable in detecting sphericity, displays a tendency towards higher Type I error rates and lower power, especially in higher-dimensional settings and smaller sample sizes. This trend is evident in its performance with exponential and lognormal distributions, where it struggles more significantly with Type I error rates and power. In contrast, Mauchly's test demonstrates a more consistent performance across different conditions, including various distributions. It maintains more stable Type I error rates and power, indicating its robustness in handling sphericity detection across different scenarios. However, both tests encounter challenges in high-dimensional cases and extreme sample sizes, which can affect their reliability. Given these limitations, incorporating the Greenhouse-Geisser correction may offer a more effective solution. This correction adjusts the degrees of freedom to account for sphericity violations, providing a more reliable adjustment across diverse sample sizes, dimensions, and distribution types, thereby improving the robustness and accuracy of statistical conclusions.

Discussion of findings

The results from the simulation study comparing Bartlett's and Mauchly's tests for sphericity under varying sample sizes (30, 50, 100, 200, 500, and 1000 observations) and dimensions (3, 5, 10, 15, 20, and 30 variables) reveal distinct patterns in performance. Bartlett's test demonstrates a relatively stable Type I error rate and power across different sample sizes and dimensions, with a slight decrease in Type I error and an increase in power as the sample size grows. This trend is consistent across dimensions, suggesting that Bartlett's test performs robustly with increasing sample sizes. In contrast, Mauchly's test exhibits a more variable performance. Although its Type I error rate decreases and power increases with larger sample sizes, it shows greater sensitivity to dimensionality, with less stability in Type I error and power in higher dimensions. Mauchly's test tends to perform well in higher dimensions but struggles with small sample sizes. Both tests face challenges with non-normal distributions, such as exponential and lognormal, where Bartlett's test consistently shows higher Type I error rates and lower power compared to Mauchly's test. Additionally, Mauchly's test, while often more sensitive to dimensionality, and provides a more reliable assessment in higher dimensions despite its susceptibility to distributional deviations.

ISSN 2736-0091
E-ISSN 2737-0107

Thus, while Bartlett's test offers stable performance with increasing sample sizes, Mauchly's test provides more reliable results in higher dimensions but is more sensitive to distributional assumptions, highlighting the importance of considering both sample size and distributional properties when choosing a sphericity test. Therefore, the choice between the two tests should be guided by the specific context of the analysis, including sample size, dimensionality, and adherence to distributional assumptions.

Conclusion

The comparison between Bartlett's and Mauchly's tests for sphericity reveals key differences in their performance based on sample size, dimensionality, and distributional assumptions. Bartlett's test shows stable performance with increasing sample sizes, demonstrating consistent Type I error rates and power. However, it struggles with non-normal distributions, exhibiting higher Type I error rates. On the other hand, Mauchly's test is more reliable in higher dimensions but is sensitive to small sample sizes and deviations from normality. Therefore, the choice of test should be informed by the specific characteristics of the data, with Bartlett's test being preferable for larger sample sizes and Mauchly's test offering advantages in high-dimensional settings. Both tests present trade-offs that should be carefully considered depending on the study's context.

Recommendations

Based on the outcomes of this study, the following recommendations are provided to guide stakeholders including researchers, data analysts, institutional policymakers, and software developers in the application and interpretation of sphericity tests in statistical analysis:

1. Researchers should adopt Bartlett's Test when working with small-to-moderate sample sizes under normal distribution conditions, due to its lower Type I error rates. For studies involving large sample sizes or high-dimensional data**,** Mauchly's Test is recommended for its relatively higher statistical power and stability.
2. Institutions and software vendors should integrate automated test selection logic into statistical tools to guide users toward the appropriate sphericity test. Additionally, capacity-building programs, including workshops, should be organized to educate researchers and graduate students on when and how to use sphericity tests, including the application of corrections in the presence of sphericity violations.
3. Academic and research institutions should implement policies that mandate the validation of sphericity assumptions before applying repeated measures ANOVA. Methodological reporting standards should also require the declaration of test type used and justification for the choice, especially in peer-reviewed research or institutional reports.
4. Stakeholders in research oversight and ethics committees should develop decision frameworks or checklists that include steps for testing statistical assumptions such as sphericity. This will ensure reproducibility and integrity in statistical reporting, especially in studies that inform public health, education, and policy decisions.

References

Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Psychology*, *3*(1), 77–85.

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems: I. Effect of inequality of variance in the one-way classification. *The Annals of Mathematical Statistics*, *25*(2), 290–302.

Efron, B., & Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*. CRC Press.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). Sage Publications.

Girden, E. R. (1992). *ANOVA: Repeated Measures*. Sage Publications.

Huynh, H., & Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational Statistics*, *1*(1), 69–82.

Jolliffe, I. T. (2002). *Principal Component Analysis* (2nd ed.). Springer.

Keselman, H. J., Algina, J., Kowalchuk, R. K., & Wolfinger, R. D. (2011). A comparison of recent approaches for the analysis of repeated measurements. *British Journal of Mathematical and Statistical Psychology*, *64*(2), 319–338.

Kovalchik, S. A., & Pfeiffer, R. M. (2014). Performing repeated measures ANOVA with missing data in SAS. *SAGE Open*, *4*(4), 2158244014559020.

Maxwell, S. E., & Delaney, H. D. (2004). *Designing Experiments and Analyzing Data: A Model Comparison Perspective* (2nd ed.). Psychology Press.

Mauchly, J. W. (1940). Significance test for sphericity of a normal n-variate distribution. *The Annals of Mathematical Statistics*, *11*(2), 204–209.

Rouder, J. N., Morey, R. D., Verhagen, J., Swagman, A. R., & Wagenmakers, E. J. (2016). Bayesian analysis of factorial designs. *Psychological Methods*, *21*(2), 229–247.

Stevens, J. P. (2009). *Applied Multivariate Statistics for the Social Sciences* (5th ed.). Routledge.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using Multivariate Statistics* (6th ed.). Pearson Education.

Wilcox, R. R. (2017). *Introduction to Robust Estimation and Hypothesis Testing* (4th ed.). Academic Press.

Zhang, X., & Chen, J. (2020). Sphericity testing in high-dimensional repeated measures designs. *Journal of Multivariate Analysis*, *179*, 104629.