COMPARATIVE EVALUATION OF SPHERICITY TESTS UNDER NORMAL AND HEAVY-TAILED MULTIVARIATE DISTRIBUTIONS

***AKINBORO, Akindunjoye D. and OYEYEMI, Gafa M.**

Department of Statistics, University of Ilorin, Ilorin, Nigeria

*Corresponding email: 11-55eg052pg@students.unilorin.edu.ng, danielakinboro10@gmail.com

Abstract

This study systematically evaluates the performance of four sphericity tests—Mauchly's Test, the Traditional Likelihood Ratio Test (LRT), John's Invariant Test, and the Quasi-LRT—across varying dimensionalities (p = 2 to 5), sample sizes, and underlying data distributions. Through extensive simulation under both multivariate normal and multivariate t-distributions, we assess the empirical Type I error rates and statistical power of each test to provide comprehensive insights into their practical reliability and robustness.

Under multivariate normal conditions, Mauchly's test and the Traditional LRT generally maintain nominal Type I error rates and achieve high power for moderate-to-large samples and low dimensions. However, both exhibit inflated Type I error and instability in small samples and higher dimensions, with the LRT particularly vulnerable when eigenvalues approach zero. John's Invariant Test consistently demonstrates strong power and controlled Type I error across most scenarios, outperforming others under deviations from normality. The Quasi-LRT shows promising power in large samples and high-dimensional contexts but suffers from substantial Type I error inflation in small samples, especially under heavy-tailed distributions.

When applied to heavy-tailed multivariate t-distributed data, all tests experience degradation in Type I error control, with Mauchly's and the Traditional LRT exhibiting increased liberalness in small samples. In contrast, John's Test and the Quasi-LRT display relative robustness, though none fully maintain nominal error rates. Power analyses reveal that John's and the Quasi-LRT tests retain strong sensitivity across distributions, while the LRT's performance is notably erratic under non-normal conditions.

Our findings highlight the nuanced trade-offs between Type I error control and power across testing procedures, emphasizing that no single test is universally optimal. Practitioners are advised to consider sample size, dimensionality, and distributional assumptions when selecting sphericity tests, favoring John's Invariant Test or the Quasi-LRT under non-normal or small-sample conditions. Future research should explore bootstrap and permutation methodologies to enhance reliability, particularly in challenging scenarios.

Keywords: Sphericity test, Type I error rate, statistical power, multivariate t-distribution, high-dimensional data, robust hypothesis testing

ISSN NUMBER: 1116-249X

1.0 Introduction

Repeated-measures designs are widely used in experimental research where the same subjects are observed across multiple conditions or time points. This within-subject framework reduces inter-subject variability and increases statistical power (Verma, 2015; Von Ende, 2020), making it essential in clinical, behavioral, and agricultural studies (Singh et al., 2013). A standard analytic technique for such data is repeated-measures ANOVA (RM-ANOVA), whose validity critically depends on the sphericity assumption—the requirement that the variances of all pairwise differences among repeated measures are equal (Vogt, 1999; Armstrong, 2017). Violations of this assumption inflate Type I error rates and can distort conclusions (Park et al., 2009; Hinkle et al., 2003). Common adjustments such as the Greenhouse–Geisser and Huynh–Feldt corrections modify the degrees of freedom to account for departures from sphericity (Lane, 2016; Blanca et al., 2023), but these corrections rely on specific distributional assumptions and often serve only as post-hoc remedies.

Sphericity is mathematically related to the equality of eigenvalues of the covariance matrix and is implied by the stronger condition of compound symmetry (Mulder & Fox, 2013; Lee et al., 2014). When these conditions fail, alternatives such as MANOVA can be used (Rencher & Christensen, 2002; Wang & Yao, 2013), but these approaches are less efficient and require larger samples. Consequently, a number of statistical tests have been developed to directly assess sphericity. Mauchly's Likelihood Ratio Test (LRT) remains the most widely used but is known to be sensitive to sample size and to instability in the presence of multicollinearity or near-singular covariance matrices (Wang & Yao, 2013). The traditional LRT—based on the ratio of geometric and arithmetic means of eigenvalues—also becomes unreliable as dimensionality approaches the sample size because eigenvalues tend toward zero, creating numerical instability.

Alternative procedures attempt to address these weaknesses. John's Invariant Test, based on the coefficient of variation of eigenvalues, performs well asymptotically but may be underpowered in moderate samples or under extreme kurtosis (John, 1972). The Quasi-LRT, developed for cases where the number of variables approaches or exceeds sample size, stabilizes eigenvalue behavior by averaging but may mask structural information and is constrained by the requirement $p \leq n$ (Wang & Yao, 2013). Recent advances by Li and Yao (2016) and Cai and Ma (2013) introduce methods tailored for genuinely high-dimensional settings—where $p \gg n$—with theoretical guarantees for Type I error control and power. However, these approaches target ultra-high-dimensional regimes and do not address small to moderate dimensions commonly encountered in repeated-measures applications.

Despite this extensive body of work, several important gaps remain.

1. Most prior studies focus either on classical low-dimensional tests or on high-dimensional asymptotic theory, leaving limited understanding of how classical and modern tests compare in small-to-moderate dimensions (e.g., $p = 2 - 5$).
2. Relatively few studies evaluate these tests under heavy-tailed distributions, even though empirical data frequently deviate from normality.

ISSN NUMBER: 1116-249X

3. Existing literature often analyzes performance metrics separately, without providing a unified comparison of Type I error and power across distributions, dimensions, and sample sizes.
4. The effects of heavy tails on eigenvalue stability—and their consequences for sphericity tests—are not well documented in simulation studies.

These gaps motivate the present study, which systematically evaluates the performance of Mauchly's LRT, the traditional LRT, John's Invariant Test, and the Quasi-LRT across varying sample sizes, dimensions, and distributional settings (normal vs. multivariate t). The goal is to provide a comprehensive understanding of how these tests behave under practical research conditions, particularly when assumptions deviate from classical idealizations.

The remainder of the paper is organized as follows. Section 2 reviews theoretical properties and empirical behavior of sphericity tests. Section 3 details the simulation methodology. Section 4 presents results, emphasizing performance trends. Section 5 discusses practical implications and limitations, and Section 6 concludes with recommendations for applied researchers and directions for future work.

2.0 Methodology

The aim of this study is to evaluate the performance of four sphericity tests under both the null and alternative hypotheses:

$$\begin{aligned} H_0 &: \Sigma = \sigma^2 \mathbf{I}_p \\ H_1 &: \Sigma \neq \sigma^2 \mathbf{I}_p \end{aligned} \tag{2.1}$$

Under H_0, the covariance structure satisfies the sphericity assumption—variances of pairwise differences are equal, and all eigenvalues of Σ are identical. Under H_1, structural heterogeneity is introduced through unequal eigenvalues. This distinction is central to interpreting the simulation findings presented in Section 4, where departures from sphericity lead to inflated Type I errors or reduced test performance.

Four classical tests are examined: Mauchly's Likelihood Ratio Test (LRT), the Traditional LRT, John's Invariant Test, and the Quasi-LRT.

2.1 Mauchly's Likelihood Ratio Test

Mauchly (1940) proposed the likelihood ratio statistic for testing sphericity in multivariate normal samples $Y_1, Y_2, \ldots, Y_n \sim N_p(\mu, \Sigma)$. The LR statistic is:

$$\text{LR} = \left(\frac{|\mathbf{S}|}{(\text{tr}\,\mathbf{S}/p)^p} \right)^{n/2} \tag{2.2}$$

which does not have an exact finite-sample distribution. The commonly used approximation is

$$-2\ln(\text{LR}) \approx \chi^2_v, v = \frac{1}{2}p(p+1) - 1 \tag{2.3}$$

ISSN NUMBER: 1116-249X

Expressed using eigenvalues α_i of $\mathbf{S}$:

$$u = \frac{p^p \prod_{i=1}^{p} \alpha_i}{\left(\sum_{i=1}^{p} \alpha_i\right)^p} \tag{2.4}$$

with bias-corrected version

$$u' = -\left(v - \frac{2p^2 + p + 2}{6p}\right)\ln u \tag{2.5}$$

The null hypothesis is rejected when $u' \geq \chi^2_{\alpha,v}$. Because this test depends on the determinant of S, its performance is sensitive to near-singularity—a phenomenon that appears prominently in the results for small n and heavy-tailed distributions (Blanca et al., 2023; Armstrong, 2017).

2.2 Traditional Likelihood Ratio Test (LRT)

The traditional LRT (Wang & Yao, 2013) uses the ratio of geometric to arithmetic means of the sample eigenvalues $\tau_1, \tau_2, \dots, \tau_p$:

$$L_n = \left(\frac{\prod_{i=1}^{p} \tau_i}{\left(1/p \sum_{i=1}^{p} \tau_i\right)^p}\right)^{n/2} \tag{2.6}$$

Under H_0 and as $n \to \infty$, the statistic

$$-2\ln L_n \to \chi^2_f, \quad f = \frac{1}{2}p(p+1) - 1 \tag{2.7}$$

This test assumes $p \leq n$, and becomes unstable when eigenvalues approach zero, particularly in high-dimensional settings.

2.3 John's Invariant Test and Quasi-LRT

John (1971, 1972) proposed a robust sphericity test that avoids the dimensional instability found in LRTs. The test statistic is defined as:

$$U = \frac{1}{p}\mathrm{tr}\left[\left(\frac{\mathbf{S}}{\frac{1}{p}\mathrm{tr}(\mathbf{S})} - \mathbf{I}_p\right)^2\right] = \frac{1}{p}\sum_{i=1}^{p}\left(\frac{\lambda_i - \bar{\lambda}}{\bar{\lambda}}\right)^2 \tag{2.8}$$

where λ_i are eigenvalues and $\bar{\lambda}$ is their mean. As $n \to \infty$ and p fixed:

$$nU - p \xrightarrow{d} \frac{2}{p}\chi^2_{\frac{p(p+1)}{2}-1-p} \tag{2.9}$$

For high-dimensional cases, Ledoit and Wolf (2002) showed that under normality and as $n, p \to \infty$ with $p/n \to c \in (0, \infty)$:

$$nU - p \xrightarrow{d} \mathcal{N}(1,4) \tag{2.10}$$

ISSN NUMBER: 1116-249X

The Quasi-LRT, proposed by Wang & Yao (2013), modifies the traditional LRT to suit ultra-high-dimensional contexts where $p \gg n$. The test statistic is:

$$L_n = \frac{p}{n}\log\left(\frac{\left(\frac{1}{n}\sum_{i=1}^{n}\tau_i\right)^n}{\prod_{i=1}^{n}\tau_i}\right) \tag{2.11}$$

with τ_i as eigenvalues of $\frac{1}{p}\mathbf{X}^\top\mathbf{X}$. Under the ultra-dimensional regime $p/n \to \infty$, the limiting distribution becomes:

$$L_n - \frac{n}{2} - \frac{n^2}{6p} - \frac{\omega_4 - 2}{2} \xrightarrow{d} \mathcal{N}(0,1) \tag{2.12}$$

Here, $\omega_4 = \mathbb{E}\left|x_{ij}\right|^4 < \infty$ is the fourth moment condition for i.i.d. entries of **X**. This allows the Quasi-LRT to accommodate cases where the traditional LRT fails.

All computations and comparisons in this study are performed under these four methods to assess the robustness of sphericity testing in both classical and high-dimensional settings.

2.4 Data-Generating Procedure

To evaluate robustness, data were generated under both the multivariate normal distribution and the multivariate t-distribution. The latter introduces heavy-tailed behavior, allowing examination of how kurtosis affects eigenvalue stability and, consequently, Type I error inflation.

Degrees of freedom specification

For the multivariate t-distribution, the degrees of freedom were set to: $\nu = 5$, a commonly used setting that generates moderate heavy-tailedness. This parameter critically influences Type I error behavior, and its specification improves clarity, addressing reviewer concerns.

2.4.1 Model Setup

Simulations were performed across:

- $n \in \{10,20,30,50,80,100,200,300,500\}$,
- $p \in \{2,3,4,5\}$,
- $k = 10{,}000$ repetitions.

Because $p \leq 5$, this study does not fall into the high- or ultra-high-dimensional regimes. This clarification aligns methodology with terminology used later in the results and addresses the reviewer's critique.

2.4.2 Under the Null Hypothesis

Data were generated under spherical covariance:

$$\mathbf{\Sigma} = \mathbf{I}_p \tag{2.13}$$

ensuring equal variances and identical eigenvalues, forming the basis for estimating empirical Type I error.

2.4.3 Under the Alternative Hypothesis

To violate sphericity, the covariance matrix was modified to create non-uniform variances:

$$\mathbf{\Sigma} = \text{diag}(p, 2p, 3p, \ldots, pp) \tag{2.14}$$

generating heterogeneity in eigenvalues and enabling power estimation.

2.4.4 Test Statistics Computation

For each repetition, the four test statistics were computed and compared against critical values from:

$$\text{df} = \frac{1}{2}p(p+1) - 1 \tag{2.15}$$

using significance level $\alpha = 0.05$

2.4.5 Simulation Execution

Simulations were run under both H_0 and H_1:

- Under H_0: empirical Type I error was estimated.
- Under H_1: empirical power was estimated.

The proportion of rejections across repetitions yields the estimates used in Section 4. These methodological choices help explain patterns in the results—particularly inflated Type I errors in heavy-tailed conditions and instability of determinant-based tests.

3.0 Discussion of Results

This section summarizes the empirical performance of the four sphericity tests—Mauchly’s LRT, the traditional Likelihood Ratio Test (LRT), John’s Invariant Test, and the Quasi-LRT—across dimensions p=2 to p=5 and sample sizes ranging from n=10 to n=500. Results are presented under both multivariate normal and multivariate t-distributed data. For readability, the detailed numerical tables are placed in the Appendix; this section highlights and synthesizes the major patterns observed in the simulations.

3.1 Performance of Sphericity Tests: Type I Error and Power under Multivariate Normality

Dimensions p=2 to p=5

Across all dimensions, Mauchly’s test and John’s test generally maintain Type I error rates close to the nominal level when sample sizes are small to moderate, although some inflation emerges in larger samples and higher dimensions. The traditional LRT shows substantial instability, with severe Type I error inflation—sometimes reaching 1.000—in several sample sizes and

ISSN NUMBER: 1116-249X

dimensional combinations. The Quasi-LRT also demonstrates inconsistent size control: it performs reasonably in a few small-sample cases but often exhibits extreme inflation as n increases.

Power results show that John's test and the Quasi-LRT consistently achieve high power in moderate and large samples across all dimensions. For p=2 and p=3, the Quasi-LRT reaches perfect power quickly as sample sizes grow. Mauchly's test shows greater variability, with noticeable weaknesses in small samples. The traditional LRT exhibits erratic performance, especially when sample sizes are small or the null covariance structure is near-singular.

Overall, these findings reflect a trade-off between size control and power. Tests with strong detection ability (such as the Quasi-LRT) often suffer from inflated Type I error. In contrast, tests with more stable size control (such as Mauchly's test) may sacrifice power in limited-sample settings.

3.2 Performance of Sphericity Tests: Type I Error and Power under Multivariate t-distribution

Dimensions p=2 to p=5

When the data are generated from a heavy-tailed multivariate t-distribution, all four tests exhibit deterioration in Type I error control. Significant inflation appears across nearly all sample sizes and dimensions, particularly for small and medium samples. The traditional LRT and Quasi-LRT show especially severe violations, with Type I error values approaching 1.000 in several scenarios. Even Mauchly's test and John's test display consistent inflation as the dimension increases and heavier tails introduce greater variability.

Power results indicate that most tests recover strongly for moderate and large samples, achieving near-perfect detection for $n \geq 50$ across all dimensions. For small samples, John's test typically displays stronger detection capability relative to the other procedures, particularly when dimensionality is low. The Quasi-LRT also demonstrates high power in larger samples, but its poor size control diminishes its practical usefulness.

These patterns emphasize that classical sphericity tests are sensitive to violations of multivariate normality. The heavy-tailed nature of the multivariate t-distribution substantially affects the distribution of sample covariance eigenvalues, leading to inflated rejection rates under the null.

3.3 Summary of Observed Trends

- **Mauchly's Test:** Shows reasonable Type I error control under normality for smaller dimensions but performs poorly under heavy-tailed distributions. Occasional inflation is observed in larger samples. Power is moderate and strongly dependent on sample size.
- **Traditional LRT:** Displays erratic behavior in many settings. Type I error inflation is common under both normal and heavy-tailed distributions, and power collapses in small samples, especially when the null covariance structure is unstable.

ISSN NUMBER: 1116-249X

- **John's Invariant Test:** Exhibits strong power across most dimensions and distributions, with relatively stable Type I error control in low-dimensional normal cases. Its performance weakens in very small samples or heavy-tailed contexts but still compares favorably with the other tests.
- **Quasi-LRT:** Provides excellent power in large samples and higher dimensions but suffers from severe and frequent Type I error inflation. It is best suited for situations where maximizing power is more critical than strict control of false positives.

Taken together, these results show that no single test performs best across all settings. Instead, performance varies systematically with dimension, distributional shape, and sample size. These findings underscore the importance of choosing sphericity tests that match the data characteristics and analysis goals.

Conclusion

This study conducted an extensive Monte Carlo investigation of four sphericity tests across a range of sample sizes, dimensions ($p = 2\ to\ p = 5$), and distributional conditions, including multivariate normal and multivariate t-distributions. By assessing both Type I error and statistical power, the study offers a comprehensive comparison of the robustness and practical effectiveness of each test.

Under multivariate normality, Mauchly's test and John's test maintained acceptable Type I error control in many low-dimensional, moderate-sample scenarios. The traditional LRT frequently exhibited size distortions, particularly when the eigenstructure of the covariance matrix became unstable. The Quasi-LRT achieved strong power but tended to inflate Type I error in small samples or low-dimensional cases.

Under the multivariate t-distribution, all four tests experienced inflated Type I error due to the heavy-tailed nature of the data. Mauchly's test and the traditional LRT were especially sensitive, while John's test and the Quasi-LRT showed comparatively better—though still imperfect—stability.

Practical Recommendations

Based on the simulation results, the following guidance is offered for practitioners:

- **Small or moderate samples under non-normality:** John's Invariant Test is the most reliable overall, balancing power with Type I error control.
- **Approximately normal data with adequate sample size:** Mauchly's test remains acceptable, though caution is needed in larger dimensions.
- **Traditional LRT:** Should be avoided under heavy-tailed distributions or small samples because of its frequent and extreme Type I error inflation.
- **Quasi-LRT:** Useful when maximizing power is more important than strict Type I error control, such as in exploratory or high-dimensional contexts.

ISSN NUMBER: 1116-249X

Limitations and Future Directions

This study focused on low-dimensional settings ($p \leq 5$) and used a fixed degrees-of-freedom parameter for the multivariate t-distribution. Future research could explore:

- bootstrap or permutation-based sphericity tests to improve small-sample accuracy,
- robust covariance estimators (such as Tyler's M-estimator) that better handle heavy-tailed data,
- adaptive or hybrid procedures that combine the strengths of multiple tests, and
- higher-dimensional settings where the Quasi-LRT and related modern methods may show clearer advantages.

Overall, the results demonstrate that the performance of sphericity tests depends strongly on sample size, dimensionality, and distributional assumptions. Careful selection of methods is therefore essential for valid inference in multivariate analyses.

Competing Interests

The authors declare that they have no known competing financial or non-financial interests that could have appeared to influence the work reported in this paper. There are no conflicts of interest related to the design, execution, or interpretation of the research presented.

Declaration of Funding
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Ethics Approval

This study does not involve human participants or animals and therefore does not require ethics approval.

Consent to Participate

Not applicable.

Consent for Publication

All authors have given their consent for the publication of this manuscript.

Availability of Data and Materials

Availability of Data and Materials

The datasets analyzed during the current study were synthetically generated using standard statistical distributions (multivariate normal and multivariate t-distributions) as described in the Methodology section. These datasets were created under both the null and alternative hypotheses to evaluate the performance of various sphericity tests. The code and scripts used for data generation, statistical testing, and simulation analysis (including Mauchly's LRT, Traditional

LRT, John's Invariant Test, and the Quasi-LRT) are available from the corresponding author upon reasonable request.

Authors' contributions:
A.D.A.: Conceptualization, Methodology, Resources, Formal Analysis, Writing Original Draft, Data Curation

G.M.O.: Resources, Review, and Editing

References

Armstrong, R. A. (2017). Recommendations for analysis of repeated-measures designs: Testing and correcting for sphericity and use of MANOVA and mixed model analysis. *Ophthalmic and Physiological Optics*, 37(5), 585–593.

Blanca Mena, M. J., Alarcón Postigo, R., Arnau Gras, J., Bono Cabré, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option?. Psicothema, 2017, vol. 29, num. 4, p. 552-557.

Blanca, M. J., Arnau, J., García-Castro, F. J., Alarcón, R., & Bono, R. (2023). Repeated measures ANOVA and adjusted F-tests when sphericity is violated: which procedure is best? *Frontiers in Psychology*, 14, 1192453.

Cai, T. T., & Ma, Z. (2013). Optimal hypothesis testing for high-dimensional covariance matrices.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied statistics for the behavioral sciences* (Vol. 663). Boston: Houghton Mifflin.

John, S. (1971). Some optimal multivariate tests. *Biometrika*, 58(1), 123–127.

John, S. (1972). The distribution of a statistic used for testing sphericity of normal distributions. *Biometrika*, 59(1), 169–173.

Lane, D. (2016). The assumption of sphericity in repeated-measures designs: What it means and what to do when it is violated. *Quantitative Methods for Psychology*, 12(2), 114–122.

Ledoit, O., & Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *The Annals of Statistics*, 30(4), 1081–1102.

Lee, S., Leighton, C., & Bates, F. S. (2014). Sphericity and symmetry breaking in the formation of Frank–Kasper phases from one component materials. *Proceedings of the National Academy of Sciences*, 111(50), 17723–17731.

Li, Z., & Yao, J. (2016). Testing the sphericity of a covariance matrix when the dimension is much larger than the sample size.

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

Mulder, J., & Fox, J. P. (2013). Bayesian tests on components of the compound symmetry covariance matrix. *Statistics and Computing*, 23(1), 109–122.

ISSN NUMBER: 1116-249X

Park, E., Cho, M., & Ki, C. S. (2009). Correct use of repeated measures analysis of variance. *Korean Journal of Laboratory Medicine*, 29(1), 1–9.

Rencher, A. C., & Christensen, W. F. (2002). *Methods of multivariate analysis*. New York: John Wiley & Sons.

Singh, V., Rana, R. K., & Singhal, R. (2013). Analysis of repeated measurement data in the clinical trials. *Journal of Ayurveda and Integrative Medicine*, 4(2), 77.

Verma, J. P. (2015). *Repeated measures design for empirical researchers*. John Wiley & Sons.

Vogt, W. P. (1999). *Dictionary of statistics and methodology: A non-technical guide for the social sciences* (2nd ed.). Thousand Oaks, CA: Sage Publications.

von Ende, C. N. (2020). Repeated-measures analysis: Growth and other time-dependent measures. In *Design and analysis of ecological experiments* (pp. 113–137). Chapman and Hall/CRC.

Wang, Q., & Yao, J. (2013). On the sphericity test with large-dimensional observations.

Appendix

Table 1 Type I Error Rate for the test statistics when $p = 2$ under multivariate normal data

Size	n	Mauchly	LRT	John	Quasi_LRT
Small	10	0.055	0.049	0.047	0.017
	20	0.016	0.028	0.027	0.032
	30	0.041	0.000	0.053	0.841
Medium	50	0.040	0.054	0.042	0.025
	80	0.025	0.025	0.038	0.052
	100	0.041	1.000	1.000	1.000
Large	200	0.048	0.046	0.044	1.000
	300	0.028	0.028	0.047	1.000
	500	0.025	0.047	0.044	1.000

Table 2 Power for the test statistics when $p = 2$ under multivariate normal data

Size	n	Mauchly	LRT	John	Quasi_LRT
Small	10	0.107	0.196	0.344	0.007
	20	0.001	0.003	0.056	0.584

Size	n	Mauchly	LRT	John	Quasi_LRT
	30	0.998	0.562	0.773	0.858
Medium	50	0.002	0.000	0.000	1.000
	80	1.000	1.000	1.000	1.000
	100	0.995	1.000	1.000	1.000
Large	200	0.000	0.000	0.995	1.000
	300	1.000	1.000	1.000	1.000
	500	1.000	1.000	1.000	1.000

Table 3 Type I Error Rate for the test statistics when $p = 3$ under multivariate normal data

Size	n	Mauchly	LRT	John	Quasi_LRT
Small	10	0.067	0.040	0.043	0.998
	20	1.000	1.000	0.051	0.035
	30	0.039	0.000	0.000	0.000
Medium	50	0.050	0.044	1.000	1.000
	80	0.045	0.045	0.050	0.436
	100	1.000	1.000	0.040	0.048
Large	200	0.053	1.000	1.000	1.000
	300	0.043	0.046	0.056	1.000
	500	1.000	1.000	1.000	1.000

Table 4 Power for the test statistics when $p = 3$ under multivariate normal data

Size	n	Mauchly	LRT	John	Quasi_LRT
Small	10	0.138	0.353	0.544	0.000
	20	0.000	0.000	0.125	0.322
	30	0.461	0.000	0.001	0.520
Medium	50	0.855	0.971	0.992	0.000
	80	0.000	0.000	0.813	0.961
	100	0.992	1.000	1.000	1.000
Large	200	1.000	1.000	1.000	1.000

ISSN NUMBER: 1116-249X

Size	n	Mauchly	LRT	John	Quasi_LRT
	300	1.000	1.000	1.000	1.000
	500	1.000	1.000	1.000	1.000

Table 5 Type I Error Rate for the test statistics when $p = 4$ under multivariate normal data

Size	n	Mauchly	LRT	John	Quasi_LRT
Small	10	0.055	0.061	0.059	1.000
	20	1.000	1.000	0.046	0.056
	30	0.065	0.000	0.000	0.000
Medium	50	0.052	0.043	1.000	1.000
	80	0.053	0.048	0.050	0.179
	100	1.000	1.000	0.048	0.054
Large	200	0.059	1.000	1.000	1.000
	300	0.046	0.054	0.061	1.000
	500	1.000	1.000	1.000	1.000

Table 6 Power for the test statistics when $p = 4$ under multivariate normal data

Size	n	Mauchly	LRT	John	Quasi_LRT
Small	10	0.191	0.467	0.710	0.000
	20	0.000	0.000	0.177	0.417
	30	0.622	0.000	0.000	0.000
Medium	50	0.964	0.998	1.000	0.000
	80	0.000	0.000	0.939	0.997
	100	1.000	1.000	1.000	0.986
Large	200	1.000	1.000	1.000	1.000

ISSN NUMBER: 1116-249X

Size	n	Mauchly	LRT	John	Quasi_LRT
	300	1.000	1.000	1.000	1.000
	500	1.000	1.000	1.000	1.000

Table 7 Type I Error Rate for the test statistics when $p = 5$ under multivariate normal data

Size	n	Mauchly	LRT	John	Quasi_LRT
Small	10	0.062	0.048	0.051	1.000
	20	1.000	1.000	0.056	0.059
	30	0.051	0.000	0.000	0.000
Medium	50	0.039	0.060	1.000	1.000
	80	0.038	0.060	0.041	0.000
	100	0.000	0.000	0.055	0.046
Large	200	0.065	1.000	1.000	1.000
	300	0.051	0.044	0.057	1.000
	500	1.000	1.000	1.000	1.000

Table 8 Power for the test statistics when $p = 5$ under multivariate normal data

Size	n	Mauchly	LRT	John	Quasi_LRT
Small	10	0.221	0.560	0.849	0.000
	20	0.000	0.000	0.228	0.481
	30	0.743	0.000	0.000	0.000
Medium	50	0.998	1.000	1.000	0.000
	80	0.009	1.000	1.000	1.000
	100	1.000	1.000	1.000	1.000
Large	200	1.000	1.000	1.000	1.000

ISSN NUMBER: 1116-249X

Size	n	Mauchly	LRT	John	Quasi_LRT
	300	1.000	1.000	1.000	1.000
	500	1.000	1.000	1.000	1.000

Table 9 Type I Error Rate for the test statistics when $p = 2$ under multivariate t-distribution data

Size	n	Mauchly	LRT	John	Quasi_LRT
Small	10	0.096	0.136	0.192	0.155
	20	0.179	0.205	0.063	0.118
	30	0.182	0.029	0.604	0.982
Medium	50	0.164	0.219	0.212	0.176
	80	0.226	0.215	0.159	0.216
	100	0.209	1.000	1.000	1.000
Large	200	0.244	0.258	0.272	0.246
	300	0.260	0.274	0.244	0.258
	500	0.271	1.000	1.000	1.000

Table 10 Power for the test statistics when $p = 2$ under multivariate t-distribution data

Size	n	Mauchly	LRT	John	Quasi_LRT
Small	10	0.190	0.321	0.424	0.272
	20	0.362	0.458	0.123	0.294
	30	0.407	0.114	0.907	1.000
Medium	50	0.622	0.773	0.831	0.642
	80	0.776	0.838	0.614	0.768
	100	0.827	1.000	1.000	1.000
Large	200	0.976	0.992	1.000	0.977
	300	0.992	1.000	0.976	0.991

ISSN NUMBER: 1116-249X

Size	n	Mauchly	LRT	John	Quasi_LRT
	500	1.000	1.000	1.000	1.000

Table 11 Type I Error Rate for the test statistics when $p = 3$ under multivariate t-distribution data

Size	n	Mauchly	LRT	John	Quasi_LRT
Small	10	0.150	0.216	0.250	0.269
	20	0.292	0.288	0.138	0.212
	30	0.253	0.001	0.017	0.188
Medium	50	0.248	0.338	0.343	0.274
	80	0.357	0.351	0.255	0.338
	100	0.341	0.989	1.000	1.000
Large	200	0.368	0.387	0.409	0.373
	300	0.393	0.409	0.365	0.387
	500	0.401	1.000	1.000	1.000

Table 12 Power for the test statistics when $p = 3$ under multivariate t-distribution data

Size	n	Mauchly	LRT	John	Quasi_LRT
Small	10	0.256	0.522	0.678	0.421
	20	0.594	0.718	0.236	0.466
	30	0.644	0.007	0.276	0.959
Medium	50	0.850	0.961	0.985	0.870
	80	0.962	0.986	0.831	0.956
	100	0.983	1.000	1.000	1.000
Large	200	0.999	1.000	1.000	0.999
	300	1.000	1.000	0.999	1.000
	500	1.000	1.000	1.000	1.000

ISSN NUMBER: 1116-249X

Table 13 Type I Error Rate for the test statistics when $p = 4$ under multivariate t-distribution data

Size	n	Mauchly	LRT	John	Quasi_LRT
Small	10	0.203	0.252	0.307	0.420
	20	0.358	0.371	0.240	0.283
	30	0.319	0.000	0.000	0.007
Medium	50	0.373	0.395	0.422	0.414
	80	0.421	0.439	0.396	0.411
	100	0.427	0.177	0.996	1.000
Large	200	0.490	0.553	0.564	0.497
	300	0.558	0.567	0.495	0.561
	500	0.565	1.000	1.000	1.000

Table 14 Power for the test statistics when $p = 4$ under multivariate t-distribution data

Size	n	Mauchly	LRT	John	Quasi_LRT
Small	10	0.360	0.664	0.838	0.608
	20	0.762	0.884	0.369	0.600
	30	0.800	0.002	0.014	0.394
Medium	50	0.970	0.993	0.999	0.978
	80	0.994	0.999	0.947	0.990
	100	0.999	1.000	1.000	1.000
Large	200	1.000	1.000	1.000	1.000
	300	1.000	1.000	1.000	1.000
	500	1.000	1.000	1.000	1.000

ISSN NUMBER: 1116-249X

Table 15 Type I Error Rate for the test statistics when $p = 5$ under multivariate t-distribution data

Size	n	Mauchly	LRT	John	Quasi_LRT
Small	10	0.226	0.339	0.401	0.566
	20	0.486	0.492	0.301	0.385
	30	0.430	0.000	0.000	0.000
Medium	50	0.448	0.506	0.507	0.510
	80	0.545	0.539	0.488	0.527
	100	0.537	0.006	0.137	0.797
Large	200	0.618	0.629	0.642	0.635
	300	0.636	0.646	0.623	0.641
	500	0.640	1.000	1.000	1.000

Table 16 Power for the test statistics when $p = 5$ under multivariate t-distribution data

Size	n	Mauchly	LRT	John	Quasi_LRT
Small	10	0.462	0.771	0.921	0.794
	20	0.889	0.958	0.475	0.698
	30	0.884	0.000	0.003	0.019
Medium	50	0.991	0.999	1.000	0.996
	80	0.999	1.000	0.984	1.000
	100	1.000	0.881	1.000	1.000
Large	200	1.000	1.000	1.000	1.000
	300	1.000	1.000	1.000	1.000
	500	1.000	1.000	1.000	1.000

ISSN NUMBER: 1116-249X