Faculty of Natural and Applied Sciences Journal of Scientific Innovations
Print ISSN: 2814-0877 e-ISSN: 2814-0923
www.fnasjournals.com
Volume 4; Issue 2; October 2023; Page No. 154-164.

COMPARATIVE ANALYSIS OF SOME APPROACHES TO MULTIVARIATE NORMALITY TEST

Victor-Edema, U.A.
Department of Mathematics/Statistics, Ignatius Ajuru University of Education, Port Harcourt, Nigeria

Corresponding author email: uyodhu.victor-edema@iaue.edu.ng

Abstract
Many common multivariate statistical techniques rely on the assumption of multivariate normality (MVN), but this assumption can often be violated in real-world data. To assess whether data deviate from MVN, various tests have been developed, including ones that use multivariate concepts like the "standardized distance method," "Mardia skewness," and "Mardia kurtosis." In this study, Monte Carlo simulations were employed to generate synthetic data to compare these three methods. Test statistics were computed for each method and then compared to the appropriate asymptotic critical values. The outcomes, indicating whether the null hypothesis was accepted or rejected in each case, were recorded and analyzed. From the result of analysis, multivariate normality was accepted at $\alpha = 0.05$ level using Standardized Distance Test and Mardia Skewness, on the other hand, Mardia Kurtosis rejected MVN at $\alpha = 0.05$ for the simulated data which is the most effective method out of the three procedures used in this work.

Keywords: Test for Normality, Multivariate Normal, Mardia Skewness, Mardia Kurtosis, Standardized Distance Method.

Introduction
Statistics is a branch of science that is needed to make decisions, and multivariate analysis plays a key role in understanding various relationships between dependent and independent events that occur in nature. Hence, the importance of statistics in predicting future events and ensuring data are used for the benefit of man. The term "multivariate" has two distinct meanings. In everyday language, it simply refers to having many things that are distinct from one another within a group, as per the Oxford Dictionary. In the realm of statistics, "multivariate" takes on a more specific definition. It pertains to a set of techniques used for analyzing multiple statistical variables simultaneously. This includes methods that go beyond the traditional two-variable analysis, like simple linear regression and correlation. Multivariate analysis is a collection of statistical methods used to examine relationships among more than two variables simultaneously. Its primary objective is to uncover intricate patterns and correlations among these variables, providing a deeper and more nuanced understanding of a particular situation compared to the limited insights gained from bivariate analysis. Multivariate analysis serves as a valuable tool for statisticians and researchers by shedding light on the reasons behind specific outcomes, ultimately enabling informed predictions and decision-making for the future. It is worth noting that the foundation of many classical or parametric multivariate analysis techniques relies on the assumption of multivariate normality (MVN). Therefore, as suggested by Andrew et al. (1973) in Mecklin and Mundform's (2004) review of methods for assessing multivariate normality, having procedures to confirm the validity of assuming normality for a given set of multivariate observations would be highly beneficial. Lynn and Alan (2018) identified that the discussion of skewness and kurtosis in textbooks is inconsistent and scant, and applications are hardly ever covered. To remedy this, they presented a clear description of kurtosis and illustrated its significant applications.

Multivariate statistics is a broad field that encompasses various forms of analysis, each with its unique goals and applications. Understanding these different approaches and how they interrelate is a key aspect of multivariate statistics. In practical terms, applying multivariate statistics to a real-world problem often involves a combination of both single-variable (univariate) and multi-variable (multivariate) analyses to gain insights into the relationships between variables and their relevance to the specific problem at hand. One of the key statistical assumptions of most multivariate procedures is that a given data is normally distributed; as such, testing the variable for normality is a key in the analytical process and the most effective method among standardized distance method, Mardia skewness, and Mardia kurtosis for the multivariate normality test.

In essence, normality refers to assessing whether a variable follows a pattern resembling the normal distribution. Consequently, when we talk about multivariate normality, we're examining whether multivariate data aligns with the characteristics of a multivariate normal distribution. This concept, first introduced by Abraham de Moivre in 1734, visualizes the multivariate normal distribution as symmetrical, three-dimensional bell-shaped curves. In this representation, the x-axis corresponds to the values of one variable, the y-axis represents the frequency or count associated with each x-value, and the z-axis signifies the values of another variable being considered.

The assumption of multivariate normality is valid when a dataset's variables exhibit a normal distribution and collectively conform to a multivariate distribution. It is important to emphasize that these tools presuppose a shared multivariate normal distribution for the data, meaning they assume the same mean (μ) and covariance (Σ) across the variables. Consequently, in many cases, assessing multivariate normality involves examining the residuals rather than the original raw data. Here are some key points to consider: First, for a dataset to meet the criteria of multivariate normality, it is essential that each variable within it follows a normal distribution. In simpler terms, if any variable doesn't adhere to a normal distribution, then the entire set of variables cannot be considered multivariate normal. Next, to assess multivariate normality, every pair of variables within the dataset must exhibit bivariate normality. This requirement entails generating scatter plots for all possible pairs of variables to examine their joint distribution. Lastly, all linear combinations are normally distributed.

There are numerous methods and visual tools available to assess whether a dataset likely comes from a normal population. One approach is to individually examine each variable for univariate normality. However, it is important to note that even if each variable appears normally distributed on its own, it does not necessarily mean that the set of variables as a whole is normally distributed. On the other hand, when a set of variables collectively conforms to multivariate normality, it implies that each variable within that set must also follow a normal distribution pattern. Thus, testing each variable for multivariate normality would be needed. Researchers have shown that data that are carefully and effectively analyzed to elicit the vital information contained in them (such as checking if the data came from the normal population) are meaningful and helpful in enabling the researcher to choose the right method of analysis, state appropriate hypotheses, select the right statistical tools, and make the right decision. Despite these numerous gains inherent in checking data for normality, many students of statistics and researchers pay little or no attention to ascertaining the normality of their data; rather, they engage in the use of statistical procedures of analysis that do not involve a normality check or their available procedures for investigating normality, or they are ignorant of the computational procedures of these methods. This research aims to address the existing gap in the investigation of normality by presenting methods carefully chosen for their ability to strike a balance between acceptable type 1 error rates and robust statistical power, outperforming other procedures across the entire distribution spectrum. It is worth noting that many tests were excluded from consideration as they rely on empirical critical values that are not easily accessible to most users. Additionally, some tests are only applicable to bivariate cases. In contrast, a newer generation of tests has emerged, designed to excel at detecting deviations from normality within specific alternative distribution scenarios. A subset of these tests caters to particular data types, such as time series data, while others are tailored to address a limited range of alternative distributions. The goal of this study was to compare the analyses of three approaches to multivariate normality tests.

This work is limited to these methods: the standardized distance (Di^2) method, the Mardia Skewness test, and Mardia Kurtosis. It will determine if the data is normally distributed and check and compare the most effective method from the three multivariate normality tests using simulated data.

For nearly two decades, there was a noticeable scarcity of methods for evaluating multivariate normality (MVN). Fortunately, this landscape has evolved significantly, and the statistical literature now offers a wealth of MVN assessment techniques. Notable among these methods are Looney (1995), Andrew et al. (1973) as cited in Mecklin and Mundform (2004), Ebner and Henze (2020), Koziol (1986) and Koziol (1993), each providing fairly comprehensive approaches to assessing multivariate normality. Most of these approaches either combine tests for univariate normality or extend traditional univariate normality tests into the multivariate realm. Furthermore, evaluating the normality of individual variables within a P-variate distribution is one way to assess

Cite this article as:
Victor-Edema, U.A. (2023). Comparative analysis of some approaches to multivariate normality test. *FNAS Journal of Scientific Innovations,4*(2), 154-164.

multivariate normality, and this is crucial as the assumption of normality underpins many parametric statistical tests. Looney (1995) discovered that deviations from normality can impact the performance of statistical procedures like multivariate analysis of variance (MANOVA), discriminant analysis, and canonical correlation to varying degrees. However, it is worth noting that despite its importance, the assessment of multivariate normality is often overlooked, with some considering it more of an academic curiosity than a practical tool, as suggested by Horsewell (1990).

Multivariate analysis encompasses a wide range of statistical techniques designed to examine more than two variables simultaneously. These variables can be of various types, including numerical, categorical, or a mix of both. The primary objective is to uncover intricate patterns and correlations among these variables, offering a deeper and more nuanced understanding of a given situation compared to what can be achieved through bivariate analysis. In this context, tools like the standardized distance method, Mardia skewness, and Mardia kurtosis play a crucial role. Statisticians and researchers utilize these methods to gain insights into why specific outcomes occur. This understanding, in turn, empowers researchers to make well-informed predictions and decisions for the future. For this analysis, data comprising variables X, Y, and Z were generated using the Monte Carlo simulation method. In this dataset, X serves as the dependent variable, while Y and Z act as independent variables.

Multivariate analysis can be categorized into two main methods: dependence and independence methods. Dependence methods come into play when some variables are dependent on others. These methods are concerned with cause-and-effect relationships—specifically, whether the values of one or more independent variables can be employed to explain, describe, or forecast the values of a dependent variable. To illustrate, consider the example of predicting someone's "weight." In this case, independent variables like "height" and "age" might be used to make that prediction. In the realm of machine learning, dependency techniques are employed to construct predictive models. In these models, analysts input data and specify which variables are independent (those used for prediction) and which are dependent (the ones to be predicted). This allows the model to leverage the independent variables to make predictions about the dependent ones.

Interdependence methods in multivariate analysis serve a different purpose compared to dependence methods. In interdependence methods, there's no hierarchy where one variable depends on another; instead, the focus is on unravelling the structural composition and inherent patterns within a dataset. These methods aren't concerned with establishing causal relationships; rather, they aim to provide meaning to a group of variables or organize them in a meaningful manner. While dependence methods explore how certain variables influence others, independent multivariate analysis is more about comprehending the overall structure of the dataset. As an example, Kankainen et al. (2007) introduced an invariant test for multivariate normality. This test is based on the Mahalanobis distance, which measures skewness, and the matrix estimate, which gauges kurtosis, between two or more multivariate location vector estimates. It is a way to assess the distribution characteristics of the data without getting into causal relationships between variables.

Kurtosis is a statistical measure that tells us whether a distribution is more peaked (positive kurtosis) or flatter (negative kurtosis) compared to a normal distribution. When a distribution has low variance, it tends to be more peaked with lighter tails, while high kurtosis indicates a distribution with a very peaked centre and heavier tails. However, it is crucial to note that peakedness alone doesn't provide a full picture of kurtosis, as emphasized by Westfall (2014). Scheffe (1959), as cited in Cain et al. (2017), pointed out that kurtosis and skewness are vital indicators of how non-normality affects the typical inferences made in the analysis of variance. To better understand normality, researchers have devised intuitive methods like kurtosis, the standardized distance method, and skewness. Enomoto et al. (2020) proposed a transformation statistic that normalizes Mardia's multivariate kurtosis. They used Monte Carlo simulations to assess the accuracy of approximating this statistic, considering its various statistical properties such as expectation, variance, skewness, kurtosis, sample error, Type I error rate, and statistical power under different alternative distributions. Additionally, Elbami and Mukerjee (2009) explored the concept of peakedness, which does not require strict normality assumptions. They investigated the comparison of dispersions in two symmetric continuous random variables around their respective medians and also provided theoretical findings, estimators for distribution functions under symmetry and peakedness constraints, and demonstrated their consistency and weak convergence. These findings were then compared with empirical estimators, offering formulas for statistical inferences.

Cite this article as:
Victor-Edema, U.A. (2023). Comparative analysis of some approaches to multivariate normality test. *FNAS Journal of Scientific Innovations,4*(2), 154-164.

In a pivotal research paper authored by Mardia in 1970, significant advancements were put forth in the field of multivariate statistics. Mardia introduced novel metrics for quantifying multivariate skewness and kurtosis, and he subsequently devised a test for assessing multivariate normality using these metrics. This test relied on establishing the correlation between these newly introduced measures. As a result, this approach became known as the "standardized distance method." Thus, the standardized distance method is denoted as

$$\boldsymbol{D^{2i}} = (\boldsymbol{y^i} - \boldsymbol{\bar{Y}})^{\boldsymbol{1}}\boldsymbol{S^{-1}}(\boldsymbol{y^i} - \boldsymbol{\bar{Y}}) \qquad (1)$$

is a useful statistic as it provides a single summary measure of feature distribution around their centres. The standardized distance method finds wide application in the field of multivariate statistics; it takes the correlation between variables. It is also one of the most suitable methods used in multivariate normality tests which are calculated with a P-degree of freedom. The paper also introduced multivariate measures of Skewness and kurtosis.
The Mardia skewness is a metric used to gauge how a probability distribution of a real-valued random variable deviates from being symmetrical around its mean. This measure also comes with an associated p-value, providing additional insights into the distribution's asymmetry. It also measures the deviation from a normal distribution. It is denoted as

$$\text{Skm}=\frac{\boldsymbol{1}}{\boldsymbol{T}^2}\boldsymbol{\Sigma\Sigma d^3 st} \qquad (2)$$

where SKm = Skewness
dst = the element of matrix D
$\boldsymbol{\Sigma}$ = Population Covariance matrix
$\boldsymbol{\Sigma_{ii}}$ = Individual Population Covariance
Mardia Kurtosis is a statistic that measures the extent to which a distribution contains outliers and for testing multivariate normality and its corresponding p-value. It is denoted as

$$\text{Kum} =\frac{\boldsymbol{1}}{\boldsymbol{T}}\boldsymbol{\Sigma\Sigma d^3 t} \qquad (3)$$

where Kum = Kurtosis
dtt = Element of matrix D=(dtt)

For the multivariate normal distribution SKm = 0 and Kum = P (P + 2). Ward (1988) highlighted that these methods rank as some of the most effective tools for evaluating multivariate normality. He emphasized that it is practically unthinkable to conduct a thorough assessment of the performance of tests for multivariate normality without incorporating Mardia's Skewness and Kurtosis tests.

Various authors, including Balanda and MacGillivary (1988) and Rupert (1987), have provided similar explanations for kurtosis. They argue that for a group of variables to adhere to multivariate normality, each variable within that group must follow a normal distribution. However, it is important to note that even if all variables are individually normally distributed, it doesn't guarantee that the entire set of variables will collectively conform to multivariate normality. This implies that simply testing each variable for univariate normality is insufficient. To address this, Mardia (1970) introduced a test for multivariate normality, which relies on sample-based metrics of multivariate skewness and kurtosis.

Both the Mardia skewness and Mardia kurtosis are functions of the squared Mahalanobis distances which are used to detect outliers in multivariate. The outliers are the variability in the observation. It may be used to indicate an experimental error. This fact makes Mardia skewness and Mardia kurtosis, particularly the kurtosis measures useful in multivariate outliers' detection.

Materials and Methods

Statistics forms the fundamental foundation for making logical, scientifically sound decisions. Therefore, in any informed empirical decision-making process, it is imperative to gather, structure, present, and analyze numerical data using suitable methods. In this study, the data collection method employed relied on generating data through computer-based processes.

In this study, a sample size of n = 30 was carefully selected using simulation. This sample was then utilized to

conduct an analysis and comparison, assessing whether the data conforms to a normal distribution or not. Furthermore, the study aimed to ascertain which of the three methods proved most effective for conducting a multivariate normality test.
Effective data collection is crucial in statistical analysis, as errors or biases in the data collection process can have a profound impact on the outcomes. In this study, the data utilized is synthetically generated using computer-based simulations to ensure its accuracy and reliability.

Method of Analysis
The generated data are compared using three procedures namely:

- Standardized distance method (Di)
- Mardia Skewness test
- Mardia Kurtosis test

Standardized Distance Method (DI)

$$\boldsymbol{D_{2i}} = (\boldsymbol{y^i} - \boldsymbol{\bar{Y}})^{\boldsymbol{1}}\boldsymbol{S^{-1}}(\boldsymbol{y^i} - \boldsymbol{\bar{Y}}) \quad (4)$$

$$\boldsymbol{S^{-1}}\begin{bmatrix} \boldsymbol{S_{11}} & \boldsymbol{S_{12}} & \boldsymbol{S_{13}} \\ \boldsymbol{S_{21}} & \boldsymbol{S_{22}} & \boldsymbol{S_{23}} \\ \boldsymbol{S_{31}} & \boldsymbol{S_{32}} & \boldsymbol{S_{33}} \end{bmatrix}^{-1}$$

$\boldsymbol{S_{ii}}$ = Individual Sample Covariance

$$\boldsymbol{S_{ii}} = \frac{\boldsymbol{1}}{\boldsymbol{n-1}}\sum_{\boldsymbol{T=1}}^{\boldsymbol{T}}(\boldsymbol{y^i} - \boldsymbol{\bar{y^i}})^{\boldsymbol{1}}\boldsymbol{S^{-1}}(\boldsymbol{y^i} - \boldsymbol{\bar{y}})^{\boldsymbol{1}}$$

where S = Sample Covariance matrix

Mardia Skewness Test

$$\text{Skm} = \frac{\boldsymbol{1}}{\boldsymbol{T^2}}\sum_{t=1}^{T} \boldsymbol{d^3 st} \quad (5)$$

$$\boldsymbol{d_{st}} = (\boldsymbol{y^i} - \boldsymbol{\bar{y}})^{\boldsymbol{1}}\boldsymbol{\Sigma^{-1}}(\boldsymbol{y^i} - \boldsymbol{\bar{y}})$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma_{11}} & \boldsymbol{\Sigma_{12}} & \boldsymbol{\Sigma_{13}} \\ \boldsymbol{\Sigma_{21}} & \boldsymbol{\Sigma_{22}} & \boldsymbol{\Sigma_{23}} \\ \boldsymbol{\Sigma_{31}} & \boldsymbol{\Sigma_{32}} & \boldsymbol{\Sigma_{33}} \end{bmatrix}^{-1}$$

$$\boldsymbol{\Sigma_{ii}} = \boldsymbol{\Sigma_{ii}} = \frac{\boldsymbol{1}}{\boldsymbol{n}}\sum_{\boldsymbol{t=1}}^{\boldsymbol{T}}(\boldsymbol{y^i} - \boldsymbol{\bar{y^i}})^{\boldsymbol{1}}\boldsymbol{S^{-1}}(\boldsymbol{y^i} - \boldsymbol{\bar{y}})^{\boldsymbol{1}}$$

where SKm = Skewness

Dst = the element of matric D = (dst)

$\boldsymbol{\Sigma}$ = population Covariance Matrix

$\boldsymbol{\Sigma_{ii}}$ = Individual Population Covariance

Mardia Kurtosis Test

$$\text{KUm} = \sum_{\boldsymbol{t=1}} \boldsymbol{d^t tt} \quad (6)$$

where, Kum = Kurtosis

$$\textbf{dtt} = (\boldsymbol{y^i} - \boldsymbol{\bar{Y}})^{\boldsymbol{1}}\boldsymbol{S^{-1}}(\boldsymbol{y^i} - \boldsymbol{\bar{Y}})$$

$$\Sigma^{-1} = \begin{bmatrix} \varepsilon_{\boldsymbol{11}} & \varepsilon_{\boldsymbol{12}} & \varepsilon_{\boldsymbol{13}} \\ \varepsilon_{\boldsymbol{21}} & \varepsilon_{\boldsymbol{22}} & \varepsilon_{\boldsymbol{23}} \\ \varepsilon_{\boldsymbol{31}} & \varepsilon_{\boldsymbol{32}} & \varepsilon_{\boldsymbol{33}} \end{bmatrix}$$

Cite this article as:
Victor-Ec T U.A. (2023). Comparative analysis of some approaches to multivariate normality test. *FNAS Journal of Scientific Innovations,4*(2), 154-164.

t =

but,

$$\varepsilon \boldsymbol{ii} = \frac{\mathbf{1}}{\boldsymbol{n}}\sum (\boldsymbol{y^i} - \boldsymbol{\bar{y}^i})^{\mathbf{1}}\Sigma^{-\mathbf{1}}(\boldsymbol{y^i} - \boldsymbol{\bar{y}})^{\mathbf{1}}$$

where KUm = Kurtosis

Dtt = Element of matrix D = (dtt)

Hypotheses
H_{0A}: The data is normally distributed
H_{1A}: The data is not normally distributed
H_{0B}: The Standardized Distance Method is the most effective method for the multivariate normality test
H_{1B}: Standardized Distance method is not the most effective method for the multivariate normality test.
H_{0C}: Mardia Skewness is the most effective method for multivariate normality tests.
H_{1C}: Mardia Skewness is not the most effective method for the multivariate normality test.
H_{0D}: Mardia Kurtosis is the most effective method for the multivariate normality test.
H_{1D}: Mardia kurtosis is not the most effective method for the multivariate normality test.

Decision Rule
Standardized distance method (D^2i)
Reject Ho if matrix (D^2i) >(n), otherwise accept
Mardia skewness test
Reject Ho if SKm (n) otherwise accept
Mardia Kurtosis Test
Reject Ho if Kum>Kum(n) otherwise do not reject.

Level of Significance
The level of significance is $\boldsymbol{\alpha = 0.05}$ where p =3

Data Presentation and Analysis
Descriptive Statistics
The tables below and the results (outputs) from the data set used for this study.

Table 1: Summary of Descriptive Statistics for the Simulated Data

Variable	N Statistic	Mean Statistic	Std. Error	Std. Deviation Statistic	Variance Statistic
X	30	53.6667	5.45332	29.86906	892.161
Y	30	65.9000	4.84575	26.54125	704.438
Z	30	52.1333	5.24935	28.75189	826.671
Valid N (listwise)	30				

A sample size of n = 30 is involved with the three variables – X, Y and Z.

Multivariate Normality Test Statistics
Standardized Distance Method

$$S = \begin{bmatrix} 892.161 & 178.966 & -230.195 \\ 178.966 & 704.438 & -103.055 \\ -230.195 & -103.055 & 826.671 \end{bmatrix}$$

$$S^{-1} = \begin{bmatrix} 0.001257 & -0.00027 & 0.000316 \\ -0.00027 & 0.001505 & 0.000112 \\ 0.000316 & 0.000112 & 0.001312 \end{bmatrix}$$

$$\bar{Y} = \begin{bmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \end{bmatrix} = \begin{bmatrix} 53.6667 \\ 65.9 \\ 52.1333 \end{bmatrix}, \qquad (Y_i - \bar{Y}) = \begin{bmatrix} (x_i - \bar{x}) \\ (y_i - \bar{y}) \\ (z_i - \bar{z}) \end{bmatrix} = \begin{bmatrix} -2.001 \\ -1.7E-13 \\ 0.001 \end{bmatrix}$$

$$(Y_i - \bar{Y})' = (-2.001 \quad -1.7E-13 \quad 0.001)$$

$$D_i = (Y_i - \bar{Y})^T S^{-1} (Y_i - \bar{Y}) = 0.005$$

Table 2: Standardized distance method test

Model	B	Z	p-value	Decision (Alpha = 5%)
Standardized Distance	0.005	5.991	0.9634	Accept H_0

Table 2 shows that the Mahalanobis distance b calculated for the variables is 0.05 with a Z-value of 5.991 at P < 0.05, there was no statistically significant relationship in the normality of the data for X, Y and Z.

Mardia Multivariate Normality Method

Table 3: Univariate skewness and kurtosis by Mardia Method

Variable	Skewness	SE_skew	Z_skew	Kurtosis	SE_kurt	Z_kurt
X	0.052	0.427	0.121	-1.392	0.833	-1.671
Y	-0.744	0.427	-1.742	-0.912	0.833	-1.095
Z	0.018	0.427	0.043	-1.495	0.833	-1.795

Table 3 shows that for the X data set, the skewness is 0.052 (SE = 0.427), and the kurtosis of – 1.392, for the Y data set, the skewness is -0.744 (SE = 0.427) and kurtosis of -0.912, and for Z data set, the skewness is 0.018 (SE = 0.427) and kurtosis of -1.495.

Table 4: Mardia's multivariate skewness and kurtosis

Model	B	Z	p-value	Decision(Alpha = 5%)
Skewness	1.13327	5.66635	0.84247298	Accept H_0
Kurtosis	11.44378	-1.77811	0.03538583	Reject H_0

From the Mardia multivariate analysis of the dataset, the Mahalanobis distance for skewness is 1.13327 with a Z-score of 5.66635, P = 0.8424 while the Kurtosis showed that the Mahalanobis distance of 11.44378 with a Z-score of -1.77811, P = 0.035.

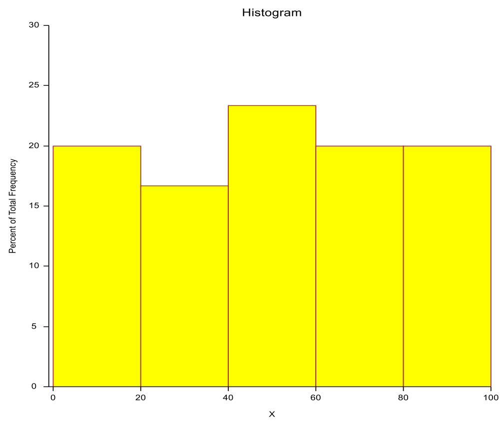
Comparison

Table 5: Comparison between the three methods

Model	B	Z	p-value	Decision(Alpha = 5%)
Standardized Distance	0.005	5.991	0.9634	Not Significant
Mardia Skewness	1.13327	5.66635	0.84247298	Not Significant

Mardia Kurtosis	11.44378	-1.77811	0.03538583	Significant

From Table 5 above, the b for the three methods of testing normality showed 0.005, 1.13327, and 11.44378 for the standardized distance method, Mardia skewness, and Mardia kurtosis respectively. At a 95% confidence interval, the P = values for the data set using the three methods are 0.9634, 0.0842 and 0.035 for standardized distance, Mardia skewness, and Mardia kurtosis respectively. The Mardia kurtosis is statistically significant for the data set while the other two are not statistically significant.



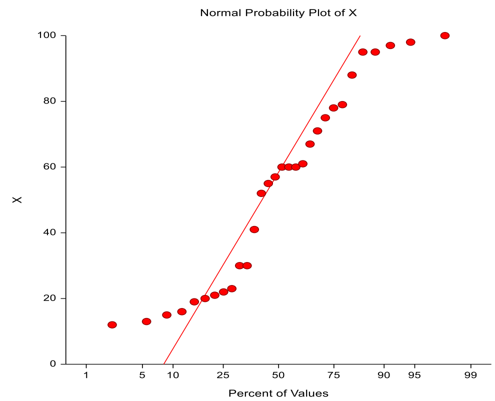
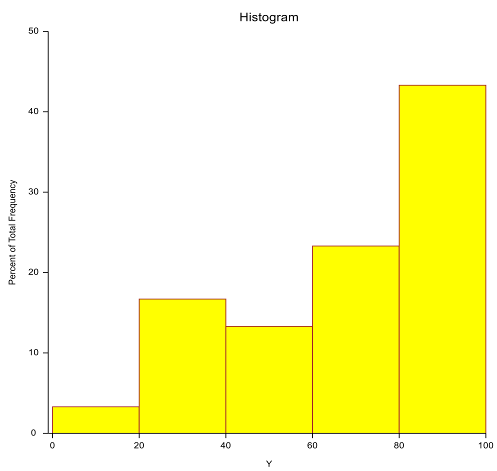


Figure 1: Histogram Plot of X

Figure 2: Normal Probability Plot of X



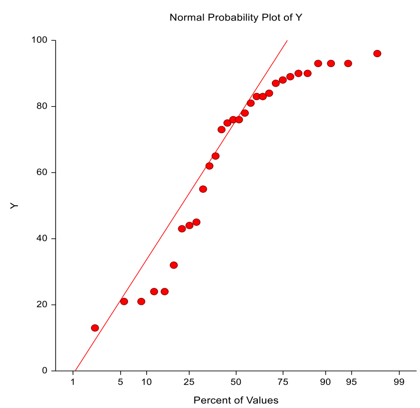


Figure 3: Histogram Plot of Y

Figure 4: Normal Probability plot of Y

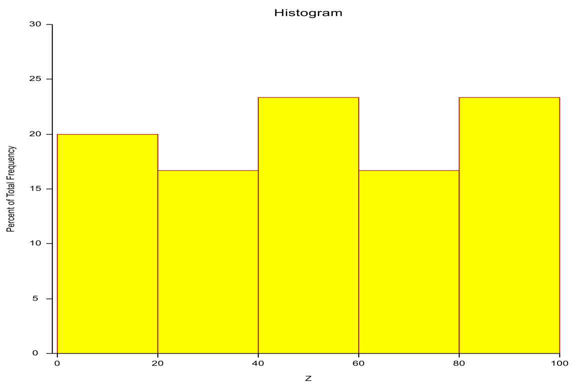


Figure 5: Histogram Plot of Z

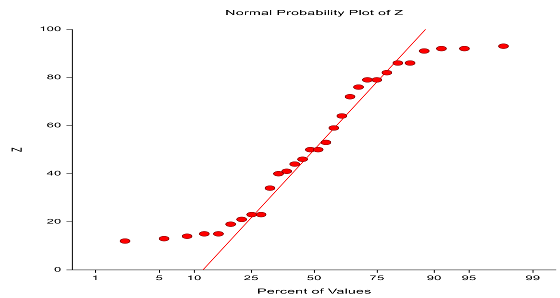


Figure 6: Normal Probability plot of Z

Discussions

Test for Normality

The result from this study revealed that the data set is relatively normal based on the fact that the skewness and kurtosis are within ± 1.95 as the standard for test for normality (Shapiro and Wilks (1965) in Górecki et al. (2018)). Furthermore, the virtual inspection of the histograms and line graphs shows some degree of normality. This finding is similar to studies cited in the literature.

Best Method for Normality Testing

From the result obtained, the Mardia kurtosis can be inferred as the best method of normality test which showed a statistically significant relationship unlike the other two tests - the standardized distance test and the Mardia Skewness which do not show a statistically significant relationship. The next effective method in the normality test is the Mardia Skewness and this is shown by a higher level of skewness expressed using the Mahalanobis distance that is greater than the value for the standardized distance approach although this value is lesser than that of the Mardia kurtosis.

It was suggested by Andrew et al (1973) in Mecklin and Mundform (2004) that there are countless possible deviations from normality as such multiple approaches for testing the multivariate normality of a data set can be reached by different procedures. Thus, from the analysis of the simulated data, multivariate normality was not rejected at the ά = 0.05 level using the standardized distance method (D_i) and Mardia skewness, on the other hand, Mardia Kurtosis rejected MVN at ά = 0.5 for the simulated data. Since the Mardia Kurtosis is significant, it is therefore concluded that the Mardia kurtosis method is the most effective method among the three methods investigated in this work.

The result from this study revealed that the Mardia kurtosis is the best method for testing for normality and this is followed by the Mardia skewness and lastly the standardized distance method. A combination of Mardia kurtosis and skewness is therefore recommended in conducting a normality test before the statistician or researcher decides on which method of analysis can be done for a given data set. Normality test is an important first-line test before carrying out any multivariate analysis and in statistics; three methods which are; the standardized distance approach, the Mardia skewness and Mardia kurtosis are commonly used. Hence, using more of the Mardia kurtosis followed by Mardia Skewness or possibly a combination of both in testing for normality as opposed to the standardized distance approach which though effective but not as effective as these two other methods.

Conclusion

The purpose of this study was to examine the multivariate normality of the generated data, comparing them and knowing the most effective ones among them using the three promising tests of multivariate normality. The test statistic for each procedure was calculated and compared with the appropriate critical value. This study analysis gave a choice on using kurtosis and Mardia skewness that they are the most useful ways in which to envision and demonstrate a departure from MVN. Indeed, Mardia skewness and kurtosis measures seem to have two serious liabilities in this regard. Kurtosis and skewness do not jointly provide a sufficient definition of MVN; that is, distribution can have MVN skewness (not skewed) and MVN kurtosis, but still be non-MVN. It is questionable whether skewness and kurtosis, at least as traditionally defined, are distinct concepts. The relevance of multivariate normality can be seen in its contribution in (i) enabling a researcher to choose the appropriate method of analysis. (ii) stating appropriate hypotheses (iii) selecting the right statistical tools and (iv) taking the right decision.

This study, therefore, proposes a prioritization of Mardia Kurtosis as the initial method for assessing multivariate data normality. Additionally, it suggests considering the use of alternative approaches like the Mardia skewness test to confirm multivariate normality, even if subsequent analyses may not strictly adhere to this assumption. To bolster the credibility of these methods, it is advisable to conduct further investigations using real-world data instead of relying solely on simulated data. Lastly, it is recommended to explore sample sizes larger than 30 to ascertain whether Mardia Kurtosis remains the most effective choice among the three methods.

References

Andrew, D.F., Gnanadesikan, R., & Warner, J. L. (1973). Methods for assessing multivariate normality: Multivariate analysis. *Proceedings of the Third International Symposium on Multivariate Analysis.* Edited by P. R. Krishnaiah, 3, 95 – 115.

Balanda, K. P., & MacGillivray, H. L. (1988). Kurtosis: A Critical Review. *The American Statistician*, *42*(2), 111–119. https://doi.org/10.2307/2684482

Cain, M.K., Zhang, Z., & Yuan, KH. (2017). Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation. *Behaviour Research Methods, 49.* 1716–1735. https://doi.org/10.3758/s13428-016-0814-1

Ebner, E. & Henze, N. (2020). Test for multivariate normality-a critical review with emphasis on weighted L_2-statistics. *Test 29*, 845-892. https://doi.org/10.1007/s11749-020-00740-0

Elbarmi, H., & Mukerjee, H. (2009). Peakedness and peakedness ordering in symmetric distributions. *Journal of multivariate analysis*, *100*(4), 594-603.

Enomoto, R., Hanusz, Z., Hara, A., & Seo, T. (2020). Multivariate normality test using normalizing transformation for Mardia's multivariate kurtosis. *Communications in Statistics - Simulation and Computation* 49(3), 684-698.

Górecki, T., Hörmann, S., Horváth, L., & Kokoszka, P. (2018). Testing normality of functional time series. *Journal of Time Series Analysis*. (JTSA) 39. 471–487.

Horswell, R. (1990) *Monte Carlo Comparison of test of Multivariate Normality Based on Multivariate Skewness and Kurtosis. PhD. Thesis*

Kankainen, A., Taskinen, S., & Oja, H. (2007). Tests of multinormality based on location vectors and scatter matrices. *Statistical Methods & Application*, *16*(3), 357-379. https://doi.org/10.1007/s10260-007-0045-9.

Koziol, J. A. (1986). An introduction to multivariate growth curve analyses. *Growth*. *50*(2), 259-72. PMID: 3792910.

Koziol, J. A. (1993). Probability Plots for Assessing Multivariate Normality. *Journal of the Royal Statistical Society. Series D (The Statistician)*, *42*(2), 161–173. https://doi.org/10.2307/2348980

Looney, S. (1995) How to use test for univariate normality. *The American Statistician 39,* 75-79.

Lynn, G. M., & Alan, F.S. (2018). Kurtosis: a forgotten moment. *International Journal of Mathematical Education in Science and Technology 49*(1), 120-130.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, *57*(3), 519-530.

Mecklin, C. J., & Mundfrom, D. (2004). An appraisal and bibliography of tests for multivariate normality. *International Statistical Review 72*(1). DOI:10.1111/j.1751-5823.2004.tb00228.x

Rupert, A. D. (1987). What is Kurtosis? An influence function approach *Journal of The American Statistical Association 41* 1–5

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika 52*: 591–611

Westfall, P. H. (2014). Kurtosis as peakedness, 1905-2014, R.I.P. *The American Statistician. 68*(3), 191-195. https://doi.org/10.1080/00031305.2014.917055

Cite this article as:
Victor-Edema, U.A. (2023). Comparative analysis of some approaches to multivariate normality test. *FNAS Journal of Scientific Innovations*,*4*(2), 154-164.

Appendix A

Table A: Stimulated Random Data

ID. Number	X	Y	Z
1	71	83	79
2	41	93	40
3	13	96	41
4	78	32	92
5	75	45	13
6	97	81	15
7	16	43	64
8	67	21	15
9	52	62	50
10	57	90	86
11	95	88	59
12	20	13	82
13	98	75	12
14	61	78	53
15	60	93	92
16	60	87	46
17	23	83	23
18	79	93	34
19	100	55	91
20	95	76	14
21	12	76	93
22	30	89	72
23	60	65	79
24	22	24	86
25	21	44	50
26	88	73	23
27	19	24	76
28	55	90	19
29	15	84	44
30	30	21	21