SciencePG
Science Publishing Group

Classification of Some Test of Normality Techniques into UMP and LMP Using Monte Carlo Simulation Technique

Awopeju Kabiru Abidemi[1, *], Ajibade Fatai Bright[2], Abuh Musa[3]

[1]Department of Statistics, Nnamdi Azikiwe University, Awka, Nigeria

[2]Petroleum Training Institute, Effurun, Warri, Nigeria

[3]Department of Statistics, Federal Polytechnic, Idah, Nigeria

Email address:

ak.awopeju@unizik.edu.ng (Awopeju Kabiru Abidemi)

[*]Corresponding author

Abstract: In Statistics, test of normality is of great importance and cannot be neglected in statistical analysis. However, there exist many techniques for such analysis and researchers usually face with the choice of test. From the literature, it has been established that power of test of normality vary significantly based on sample sizes. In this study, seven normality tests were reviewed and the classification into LMP and UMP were based on Power-of-Test. The test of hypothesis was done at 5% level of significance. The tests considered as; Shapiro-Wilk, Anderson-Darling, Bonett-Serial, Robust Jarque-Bera, Skewness, Lilliefors and Kurtosis tests. The sample sizes considered are 10, 20, 50 and 100 with 1000 replicates. Simulation was done from 3 distributions namely, normal, gamma and beta distributions. It was found that all methods were stronger for the detection of normality when normal distribution was used but the variation in their power was obvious when non-normal distributions were used. Among the methods, only three can be referred to as UMP while the rest are LMP. The UMP methods are Shapiro-Wilk, Anderson-Darling and Lilliefors as their Power-of-Test was not affected by sample sizes.

Keywords: Distributions, Simulation, Monte Carlo Techniques, Power-of-Test, Type I Error

1. Introduction

Test of normality is commonly use in data analysis irrespective of the field of study of the researchers especially in the determination of the most appropriate statistical technique for data analysis or hypothesis testing. Statistical methods were derived with some basic assumptions outside which the method becomes invalid [5]. For instance, in the case of Chi-Square test of dependency, it is assumed that the cells contain non-zero entries as observed values and more than 20% of the cells contain expected values more than 5. In a similar manner, Z-test and T-test which are often use by researchers for test of significant variation between variables are used with caution as the methods have basic assumptions such as normality of the variables and sample size of the observations [1, 25].

Generally, Statistical methods are of two types, namely, parametric and nonparametric methods. For the use of parametric methods, there is need for test of normality for the determination of suitability of the method or test statistic [20]. From the literature, there exist more than one hundred methods of test of normality for both univariate and multivariate and the sensitivity of the methods has been shown to depend on some factors such as sample size etc. According to Ukponmwan and Ajibade [35], some test of normality are more sensitive to small sample sizes and some are better used when the sample sizes are large. In the case of multivariate test of normality, sample size of the variables play key role in the choice of method of test of normality as shown in the literature [11, 13].

Despite the improvement on the normality methods as presented by researchers from different fields, there is significant or noticeable gap in the sensitivity of the methods which resulted to type I or type II error in many decision taken by the researchers. From the output of test of normality using Ryan-Joiner as presented in Minitab software and the

Kolmogorov Smirnov in a set of data, it was observed that the decision of the researcher vary which depends on the method used for the test of normality. This implies there is need for better understanding of a set of data before analysis or hypothesis testing.

The study is aimed at classification of some selected normality methods into Uniformly Most Powerful and Locally Most Powerful. For a robust study, the following objectives were considered;

1) Determination of effect of sample sizes on performance of normality test methods.
2) Rating of the methods in terms of Power-of-Test and Type-I-Error.

2. Review of Related Studies

Stephen [32] confirmed the assumption of multivariate normality (MVN). According to the researcher, for the past 50 years, over 50 tests of this assumption have been proposed. However, for various reasons, practitioners are often reluctant to address the MVN issue. In the research, several techniques for assessing MVN based on well-known tests for univariate normality were described and suggestions are offered for their practical application. The techniques are illustrated using two previously published sets of real-life data. In one of the examples it was shown that simply testing each of the marginal distributions for univariate normality can lead to a mistaken conclusion.

According to Douglas, Bonett and Edith [9], kurtosis can be measured in more than one way. A modification of Geary's measure of kurtosis is shown to be more sensitive to kurtosis in the center of the distribution while Pearson's measure of kurtosis is more sensitive to kurtosis in the tails of the distribution. The modified Geary measure and the Pearson measure are used to define a joint test of kurtosis that has high uniform power across a very wide range of symmetric non normal distributions.

Lobato, Ignacio, and Carlos [24] proposed the skewness-kurtosis test statistic, but studentized by standard error estimators that are consistent under serial dependence of the observations. The standard error estimators are sample versions of the asymptotic quantities that do not incorporate any down weighting, and, hence, no smoothing parameter is needed. Therefore, the main feature of our proposed test is its simplicity, because it does not require the selection of any user-chosen parameter such as a smoothing number or the order of an approximating model.

Thorsten and Herbert [33] confirmed various pattern of testing normality regarding to their power several tests were carried out. Well known test was Jarque and Bera [17-19], the test of Kuiper [22] and Shapiro and Wilk [29] as well as tests of Kolmogorov-Smirnov and Cramer Von Mises type. The tests on normality are based, first, on independent random variables (model I) and, second, on the residuals in the classical linear regression (model II). They investigated the exact critical values of the Jarque-Bera test and the Kolmogorov-Smirnov and Cramer Von Mises tests, in the

latter case for the original and standardized observations where the unknown parameters μ and σ have to be estimated. The power comparison is carried out via Monte Carlo simulation assuming the model of contaminated normal distributions with varying parameters μ and σ and different proportions of contamination. It turns out that for the Jarque-Bera test the approximation of critical values by the chi-square distribution does not work very well. The test is superior in power to its competitors for symmetric distributions with medium up to long tails and for slightly skewed distributions with long tails. The power of the Jarque-Bera test is poor for distributions with short tails, especially if the shape is bimodal, sometimes the test is even biased. In this case a modification of the Cram´er-von Mises test or the Shapiro-Wilk test may be recommended.

Ralph [28] viewed experimentally-derived data sets that are generated in the practice of clinical chemistry. It was stated that graphical presentation is essential to assess the data distribution. The distribution must also be assessed quantitatively. The approach determines if the data is normal or not. Finally the results of the tests of Normality must be shown to be free of sample size effects. In the work, four experimentally-derived data sets were used. They represented normal. Positive and negatively-skewed distributions were considered. These data sets were examined by graphical techniques, by moment tests, by tests of Normality, and monitored for sample size effects. In the conclusion, it was stated that the preferred graphical techniques are the histogram and the box-and-whisker plots that may be supplemented, with advantage, by quartile quartile or probability - probability plots. Classical tests of skewness and kurtosis can produce conflicting and often confusing results and. as a consequence. the alternative use often newer L-moments is advocated, Normality tests included the Kolmogorov Smirnov (Lilliefors modification), Cramer-von Mises and Anderson- Darling tests (empirical distribution function statistics) and the Gan"" Koehler. Shapiro-Wilk, Shapiro- Francia, and Filliben tests.

Shengyi, Shengyi and Robert [31] used census block groups data on socio-demographics, land use, and travel behavior, to test the cutoffs suggested in the literature for trustworthy estimates and hypothesis testing statistics, and evaluate the efficacy of deleting observations as an approach to improving multivariate normality, in structural equation modeling. The results showed that the measures of univariate and multivariate non-normality will fall into the acceptable range for trustworthy maximum likelihood estimation after a few true outliers are deleted.

Yap and Sim [36] compared the power of eight selected normality tests: the Shapiro–Wilk test, the Kolmogorov–Smirnov test, the Lilliefors test, the Cramer–von Mises test, the Anderson–Darling test, the D'Agostino–Pearson test, the Jarque–Bera test and Chi-squared test. Power comparisons of these eight tests were obtained via the Monte Carlo simulation of sample data generated from alternative distributions that follow symmetric short-tailed, symmetric long-tailed and asymmetric distributions. The simulation

results showed that for symmetric short-tailed distributions, D'Agostino and Shapiro–Wilk tests have better power. For symmetric long-tailed distributions, the power of Jarque–Bera and D'Agostino tests is quite comparable with the Shapiro–Wilk test. As for asymmetric distributions, the Shapiro–Wilk test is the most powerful test followed by the Anderson–Darling test.

Adefisoye, Golam and George [2] examined problems of testing for normality in both theoretical and empirical statistical research. The performances of eighteen normality tests were all listed. Monte Carlo simulation was used from various symmetric and asymmetric distributions for different sample sizes ranging from 10 to 100. The performance of the test statistics were compared based on empirical Type I error rate and power of the test. The simulation results showed that the Kurtosis Test is the most powerful for symmetric data and Shapiro Wilk test is the most powerful for asymmetric data.

According to Felix and Senyo [12], most parametric methods rely on the assumption of normality. Results obtained from these methods are more powerful compared to their non-parametric counterparts. However for valid inference, the assumptions underlying the use of these methods should be satisfied. According to the researchers, many published statistical articles that make use of the assumption of normality fail to guarantee it. Hence, quite a number of published statistical results are presented with errors. As a way to reduce this, in assessing univariate and multivariate normality, several methods have been proposed. In the univariate setting, the Q-Q plot, histogram, box plot, stem-and-leaf plot or dot plot are some graphical methods that can be used. Also, the properties of the normal distribution provide an alternative approach to assess normality. The Kolmogorov-Smirnov (K-S) test, Lilliefors corrected K-S test, Shapiro-Wilk test, Anderson-Darling test, Cramer-von Mises test, D'Agostinoskewness test, Anscombe-Glynn kurtosis test, D'Agostino-Pearson omnibus test, and the Jarque-Bera test are also used to test for normality. However, Kolmogorov-Smirnov (K-S) test, Shapiro-Wilk test, Anderson-Darling test, and Cramer-von Mises test are widely used in practice and implemented in many statistical applications. For multivariate normal data, marginal distribution and linear combinations should also be normal. This provides a starting point for assessing normality in the multivariate setting. A scatter plot for each pair of variables together with a Gamma plot (Chi-squared Q-Q plot) is used in assessing bivariate normality.

Ukponmwan and Ajibade [35] explained the sensitivity of nine normality test statistics; W/S, Jaque-Bera, Adjusted Jaque-Bera, D'Agostino, Shapiro-Wilk, Shapiro-Francia, Ryan-Joiner, Lilliefors'and Anderson Darlings test statistics, with a view to determining the effectiveness of the techniques to accurately determine whether a set of data is from normal distribution or not. Simulated data of sizes 5, 10, …, 100 is used for the study and each test is repeated 100 times for increased reliability. Data from normal distributions (N (2, 1) and N (0, 1)) and non-normal distributions (asymmetric and symmetric distributions: Weibull, Chi-

Square, Cauchy and t-distributions) are simulated and tested for normality using the nine normality test statistics. To ensure uniformity of results, the researchers used one software in all the data computations to eliminate variations due to statistical software. The error rate of each of the test statistic is computed; the error rate for the normal distribution is the type I error and that for non-normal distribution is type II error. Power of test is computed for the non-normal distributions and use to determine the strength of the methods. The ranking of the nine normality test statistics in order of superiority for small sample sizes is; Adjusted Jarque-Bera, Lilliefor's, D'Agostino, Ryan-Joiner, Shapiro-Francia, Shapiro-Wilk, W/S, Jarque-Bera and Anderson-Darling test statistics while for large sample sizes, we have; D'Agostino, Ryan-Joiner, Shapiro-Francia, Jarque-Bera, Anderson-Darling, Lilliefor's, Adjusted Jarque-Bera, Shapiro-Wilk and W/S test statistics. Hence, only D'Agostino test statistic is classified as Uniformly Most Powerful since it is effective for both small and large sample sizes. Other methods are Locally Most Powerful. Shapiro-Francia, an improvement of Shapiro-Wilk is more sensitive for both small and large samples, hence should replace Shapiro-Wilk while the Adjusted Jarque-Bera and the Jarque-Bera should both be retained for small and large samples respectively.

Tanweer [34] evaluates the performance of some selected normality tests for the skewed alternative space. According to the researcher, stringency concept allows rank of 12 tests uniquely. Among the methods considered, Bonett and Seier test (Tw) turns out to be the best statistics for slightly skewed 13 alternatives and the Anderson-Darling (AD), Chen-Shapiro (CS), Shapiro-Wilk (W) and Bispo, 14 Marques, &Pestana, (BCMR) statistics were the best choices for moderately skewed alternative 15 distributions. Maximum loss of Jarque-Bera (JB) and its robust form (RJB), in terms of deviations 16 from the power envelope, is greater than 50% even for large sample sizes which makes them less 17 attractive in testing the hypothesis of normality against the moderately skewed alternatives. On 18 balance, all selected normality tests except Tw and COIN performed exceptionally well against the 19 highly skewed alternative space.

According to Agnieszka [3], statistical inference in the form of hypothesis tests and confidence intervals often assumes that the underlying distribution is normal. Similarly, many signal processing techniques rely on the assumption that a stationary time series is normal. As a result, a number of tests have been proposed in the literature for detecting departures from normality. In this research Robert developed a novel approach to the problem of testing normality by constructing a statistical test based on the Edgeworth expansion, which approximates a probability distribution in terms of its cumulates. By modifying one term of the expansion, a define test statistic which includes information on the first four moments. performs a comparison of the proposed test with existing tests for normality by analyzing different platykurtic and

leptokurtic distributions including generalized Gaussian, mixed Gaussian, α-stable and Student's t distributions. Robert showed that some proposed test is superior in terms of power for the platykurtic distributions whereas for the leptokurtic ones it is close to the best tests like those of D'Agostino-Pearson, Jarque-Bera and Shapiro-Wilk.

Bruno and Norbet [6] confirmed the developments in affine invariant tests for multivariate normality with special emphasis on asymptotic properties of several classes of weighted L^2-statistics. According to the researchers, weighted L^2-statistics typically have limit normal distributions under fixed alternatives to normality, they open ground for a neighborhood of model validation for normality. The paper also reviews several other invariant tests for this problem, notably the energy test, and it presents the results of a large-scale simulation study. All tests under study are implemented in the accompanying R-package.

Nasrin [26] showed that different univariate normality testing procedures are compared by using new algorithm. Different univariate and multivariate test are also analyzed and also review efficient algorithm for calculating the size corrected power of the test which can be used to compare the efficiency of the test. In the research, 100 data sets with combinations of sample sizes; n = 10, 20, 25, 30, 40, 50, 100, were generated from uniform distribution and tested by using different tests for randomness. The assessment of normality using statistical tests is sensitive to the sample size. It was observed that with the increase of sample size, overall powers are increased but Shapiro Wilk (SW) test is the most powerful test among other tests.

3. Methodology

Every research study has the main aim of finding solution to some define problems and in an attempt to solve the problems, various techniques and strategies are employed. The research process usually involves identifying the problems, making Hypothetical statements about the pleasured relationship among the variables, collecting and analyzing such data using the appropriate statistical tools. This procedure is known as the research mythology. In this section, these processes mentioned above will be treated and also the statistical tools to be employed in the analysis of the data highlighted.

3.1. Statistical Methods

The following univariate normality tests were considered in the research;

3.1.1. Anderson-Darling Test [AD]
The AD test is of the form:

$$AD = n \int_{-\infty}^{\infty} [f_n(x) - \Phi(x)]^2 \psi(x) df(x)$$

Where $f_n[x]$ is the empirical distribution function, $\phi[x]$ is the cumulative distribution function of the standard normal distribution and $\Psi[x]$ is a weight function [4].

3.1.2. D'Agostino-Pearson K^2 test [DK]
The test combines g_1 and g_2 to produce an omnibus test of normality. The test statistics is:

$$K^2 = Z^2_s + Z^2_k$$

Z^2_s and Z^2_k are the normal approximations to sand k respectively [7]. The test statistic follows approximately a chi-square distribution with 2 degree of freedom when a population is normally distributed. The test is appropriate for a sample size of at least twenty.

3.1.3. Shapiro-Wilk Test [sw]
The Shapiro-wilk test uses a w-static [14, 30] which is defined as

$$W = \frac{1}{D} [\sum_{i=1}^{m} a_i (x_{(n-i+1)} - x_{(i)}]^2$$

Where m $= \frac{n}{2}$ if n is even while m = [n – 1]/2 if n is odd.
$D = \sum_{i=1}^{n} [x_i - \bar{x}]^2$ and $x_{[i]}$ represents the ith order statistic of the sample.

The constants a_i are given by
$(a_1, a_2, \dots, a_n) = \frac{m'v^{-1}}{(m'v^{-1}v^{-1}m)^{1/2}}$ and m is given by m = $(m_1, m_2, \dots, m_n)$ where $m_1, m_2, \dots m_n$ are the expected values of the order statistics of independent and identically distributed random variables sampled from the standard normal distribution, and V is the covariance matrix of those order statistics.

3.1.4. Bonett-Serial Test [BS]
The test statistic T_w is given by:
$T_w = \frac{\sqrt{n} + 2.[w - 3]}{3.54}$ in which w is set by

$$w = 13.29[\ln\sqrt{m_2} - \ln[n^{-1}\sum_{i=1}^{n} |x_I - \bar{x}|]]$$

This statistic follows a standard normal distribution under null hypothesis.

3.1.5. Robust Jarque-Bera Test [RJB]
The robust Jarque-Bera [RJB test statistics define as

$$RJB = \frac{n}{6} [\frac{m_3}{J_n^3}]^2 + \frac{n}{64} [\frac{m_4}{J_n^4} - 3]^2,$$

Where $J_n = \frac{c}{n}\sum_{i=1}^{n} |x_i - m|, C = \sqrt{\pi/2}$ and M is the sample median. The RJB statistic is asymptotically X_2^2-distributed.

3.1.6. Lilliefor's Test (LL)
The test statistics is defined as;

$$D = Sup_x |F^*(x) - S_n(x)|$$

where $S_n(x)$ is the sample cumulative distribution function and $F^*(x)$ is the cumulative distribution function (CDF) of the null distribution [23].

3.1.7. Cramer-Von Mises Test (CVM)
In statistics the cramer-von mises criterion is a criterion used for judging the goodness of fit of a cumulative distribution function F^* compared to a given empirical

distribution function F_n, or for comparing two empirical distributions. It is also used as a part of other algorithms, such as minimum distance estimation. It is defined as;

$$W^2 = \int_{-\infty}^{\infty} [\,F_n(x) - F^x(x)\,]^2 df^*(x)$$

3.1.8. Skewness Test [SK]

The skewness statistic is defined as

$$SK = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^3}{(\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2)^{3/2}}$$

Here, $\bar{x}$ is the sample mean.

3.1.9. Kurtosis Test [KU]

The Kurtosis statistic is defined as:

$$KU = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^4}{(\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2)^2} - 3$$

Here, $\bar{x}$ is sample mean. The " minus 3" at the end of this formula is often explained as a correction to make the kurtosis of the normal distribution equal to zero, as the kurtosis is 3 for a normal distribution.

3.2. Data Presentation

Data were stimulated from normal and non normal distributions with varying sample sizes (10, 20, 50 and 100). A replicate of 1000 was used at every point.
Distribution of interest;
1. Gamma Distribution
2. Beta Distribution
3. Uniform Distribution
4. Normal Distribution

3.3. Grouping of Normality Techniques into UMP and LMP

Normality test statistic that is capable of detecting normality for both large and small sample sizes can be referred to as UMP. Test statistic that can only be used for either small or large sample size can be referred to as LMP [10]. This implies any normality test technique classified as UMP can be used for test of normality, irrespective of sample size of the data and the LMP tests must be used with caution.

3.4. Algorithm for Monte Carlo Simulation

According to Ukponmwan and Ajibade [35], Monte Carlo (MC) simulation can be used for the performance of an estimator. Monte Carlo procedure can be stated thus;
1. Specify the data simulation procedure.
2. Select desirable sample size n.

3. Decide the desirable number of replicates.
4. Generate a random sample of size n based on the Data.
5. Using random sample generated, calculate the statistic(s).
6. Go back to (4) and repeat (4) and (5) until desirable result is achieved.
7. Examine the results for type I or type II error.

3.5. Determination of Sensitivity of Normality Test

In this paper, P-value was used for the determination of sensitivity of normality testing techniques considered. The decision rule can be stated as; reject the null hypothesis if the p-value is less than the level of significance. In this study, a p-value greater than 0.05 implies the set of data is normally distributed and the higher the value.

According to Nornadiah and Yap [27], power of a statistical test is the probability that the test will reject the null hypothesis, when the alternative hypothesis is true (i.e. the probability of not committing a Type II error). The power is in general a function of the possible distributions, often determined by a parameter, under the alternative hypothesis. As the power of test increases, the chance of a Type II error occurring decreases [16, 21]. The probability of a Type II error occurring is referred to as the false negative rate (β). Therefore power of test is equal to $1 - β$, which is also known as the sensitivity [8, 15].

4. Analysis

In this section, simulation was carried out using symmetric and asymmetric distributions to capture possible distributions in existence in Statistics. Using the selected techniques, the data analysis was done using the steps below;
i. Simulate data of sizes n = 10, 20, 50 and 100 from the four distributions
ii. For each sample size, n, replicate 100 times,
iii. Calculate the normality test statistics for each of the univariate normality tests
Ii Reject/Accept the null hypothesis;
H_0: the data are normally distributed.
H_1: the data are not normally distributed.
iv. Calculate the error rate = $\frac{number\ of\ times\ wrong\ decision\ is\ recorded}{number\ of\ replicate\,(1000)}$
v. Calculate the type I error or power of test = 1- p(type II error)

Analysis and Discussion of Result: Using R statistical package and Monte Carlo procedure, the following results were obtained;

Table 1. *Error Rate of SW.*

	Gamma distribution	Betadistribution	Uniform distribution	Normal distribution
10	0.278	0.274	0.227	0.050
20	0.169	0.104	0.221	0.052
50	0.123	0.180	0.140	0.050
100	0.093	0.088	0.032	0.052

Table 2. *Error Rate of Anderson Darling.*

	Gamma distribution	Betadistribution	Uniform distribution	Normal distribution
10	0.178	0.188	0.086	0.049
20	0.128	0.174	0.077	0.053
50	0.099	0.042	0.069	0.042
100	0.074	0.007	0.021	0.049

Table 3. *Error Rate of Bonett-Serial Test.*

	Gamma distribution	Betadistribution	Uniform distribution	Normal distribution
10	0.012	0.093	0.055	0.007
20	0.005	0.092	0.032	0.006
50	0.009	0.078	0.011	0.008
100	0.007	0.062	0.003	0.007

Table 4. *Error Rate of Robust Jarque-Bera Test.*

	Gamma distribution	Betadistribution	Uniform distribution	Normal distribution
10	0.042	0.024	0.065	0.009
20	0.082	0.109	0.058	0.016
50	0.091	0.185	0.023	0.040
100	0.126	0.365	0.022	0.053

Table 5. *Error Rate of D'Agostino-Person Test.*

	Gamma distribution	Betadistribution	Uniform distribution	Normal distribution
10	0.026	0.080	0.034	0.044
20	0.057	0.050	0.023	0.051
50	0.021	0.080	0.022	0.054
100	0.021	0.077	0.014	0.053

Table 6. *Error Rate of Skewness Test.*

	Gamma distribution	Betadistribution	Uniform distribution	Normal distribution
10	0.026	0.059	0.024	0.002
20	0.057	0.055	0.022	0.020
50	0.089	0.043	0.014	0.030
100	0.059	0.042	0.008	0.033

Table 7. *Error Rate of Kurtosis Test.*

	Gamma distribution	Betadistribution	Uniform distribution	Normal distribution
10	0.086	0.093	0.087	0.056
20	0.089	0.086	0.054	0.062
50	0.090	0.083	0.035	0.044
100	0.014	0.081	0.024	0.020

Table 8. *Error Rate of Lilliefores (LL) Test.*

	Gamma distribution	Betadistribution	Uniform distribution	Normal distribution
10	0.081	0.071	0.066	0.049
20	0.066	0.039	0.054	0.058
50	0.044	0.023	0.032	0.041
100	0.011	0.010	0.023	0.053

Table 9. *Error Rate of Cramer–von Mises (CVM) Test.*

	Gamma distribution	Betadistribution	Uniform distribution	Normal distribution
10	0.065	0.033	0.045	0.059
20	0.060	0.032	0.043	0.053
50	0.017	0.012	0.025	0.057
100	0.004	0.008	0.011	0.041

From the tables presented above, it can be observed that no normality test statistic has 100 percent accuracy with respect to

samples sizes considered as well as the distributions considered. This is a signal that univariate normality test should be used with utmost caution to prevent types of error in hypotheses testing.

Table 10. *Power-of-Test of Selected Normality Test Statistics.*

	10	20	50	100
Shapiro-Wilk				
Gamma	0.722	0.831	0.877	0.907
Beta	0.726	0.896	0.82	0.912
Uniform	0.773	0.779	0.86	0.968
Normal	0.95	0.948	0.95	0.948
Anderson Darling				
Gamma	0.822	0.872	0.901	0.926
Beta	0.812	0.826	0.958	0.993
Uniform	0.914	0.923	0.931	0.979
Normal	0.951	0.947	0.958	0.951
Bonett-Serial				
Gamma	0.988	0.995	0.991	0.993
Beta	0.907	0.908	0.922	0.938
Uniform	0.945	0.968	0.989	0.997
Normal	0.993	0.994	0.992	0.993
Robust Jarque-Bera				
Gamma	0.958	0.918	0.909	0.874
Beta	0.976	0.891	0.815	0.635
Uniform	0.935	0.942	0.977	0.978
Normal	0.991	0.984	0.960	0.947
D'Agostino-Person				
Gamma	0.974	0.943	0.979	0.979
Beta	0.920	0.950	0.920	0.923
Uniform	0.966	0.977	0.978	0.986
Normal	0.956	0.949	0.946	0.947
Skewness				
Gamma	0.974	0.943	0.911	0.941
Beta	0.941	0.945	0.957	0.958
Uniform	0.976	0.978	0.986	0.992
Normal	0.998	0.980	0.970	0.967
Kurtosis				
Gamma	0.914	0.911	0.910	0.986
Beta	0.907	0.914	0.917	0.919
Uniform	0.913	0.946	0.965	0.976
Normal	0.944	0.938	0.956	0.980
Lilliefores (LL)				
Gamma	0.919	0.934	0.956	0.989
Beta	0.929	0.961	0.977	0.990
Uniform	0.934	0.946	0.968	0.977
Normal	0.951	0.942	0.959	0.947
Cramer–von Mises (CVM)				
Gamma	0.935	0.940	0.983	0.996
Beta	0.967	0.968	0.988	0.992
Uniform	0.955	0.957	0.975	0.989
Normal	0.941	0.947	0.943	0.959

Table 10 shows strength of univariate normality tests considered with respect to some selected distributions. From the table, it can be observed that power of univariate test statistic is significantly affected by nature of the data (distribution). At this point, it is adequate to conclude that sample size and shape of data play vital role in the choice of normality test.
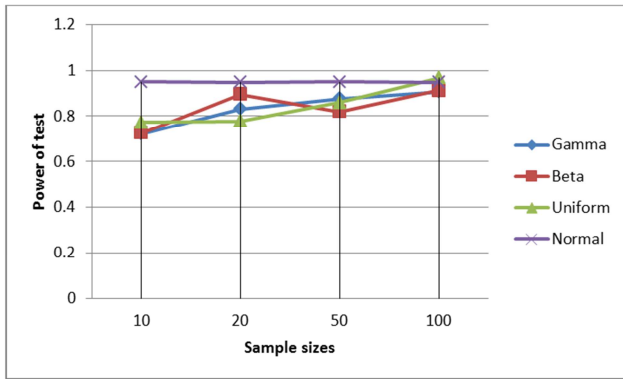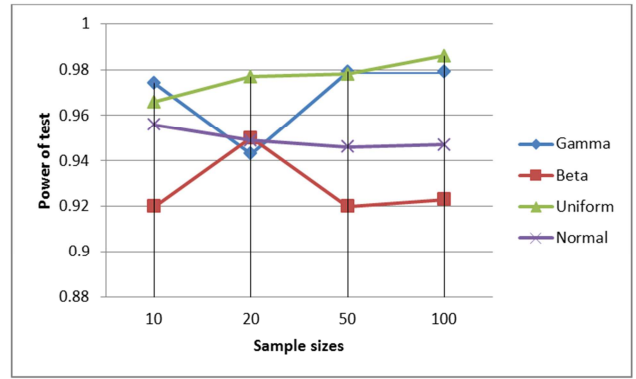
Figure 1. Power of test of Shapiro-Wilk test.
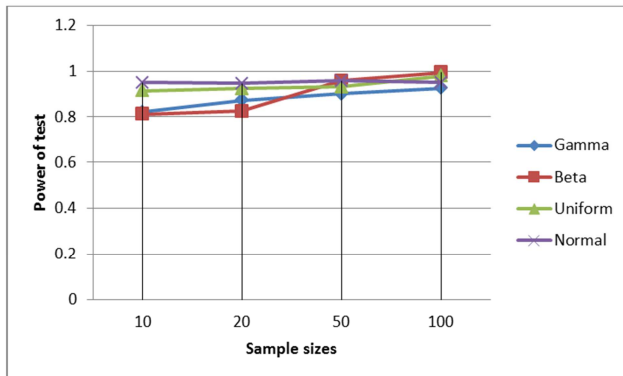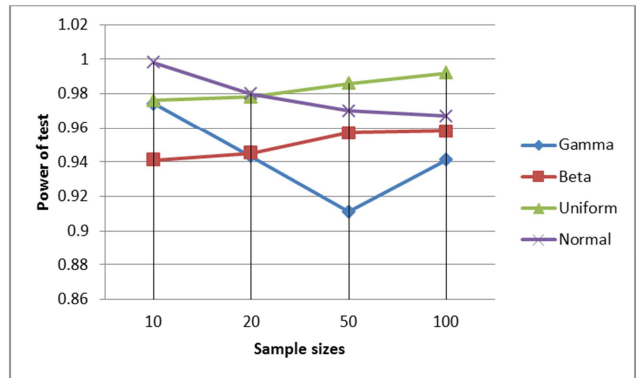

Figure 2. Power of Test of Anderson Darling test.

Figure 3. Power of test of Bonett-Serial test.

Figure 4. Power of test of Robust Jarque-Beratest.

Figure 5. Power of test of D'Agostino-Person test.
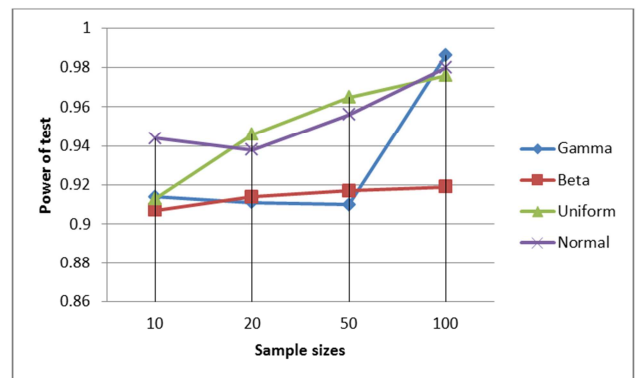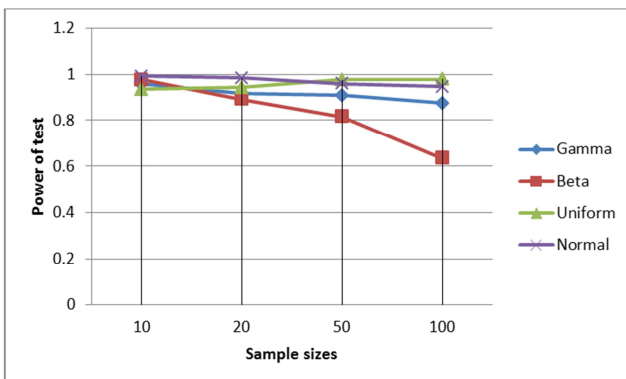

Figure 6. Power of test ofSkewness test.

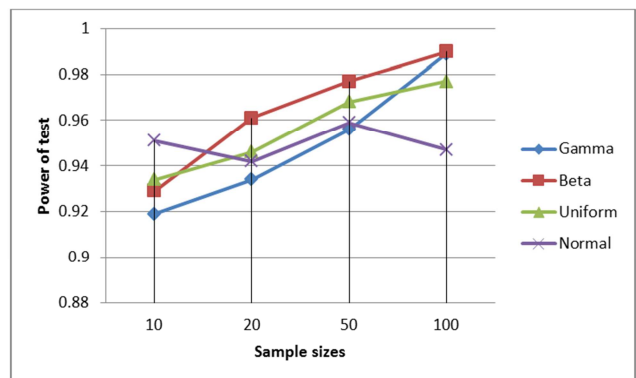Figure 7. Power of test of Kurtosis test.

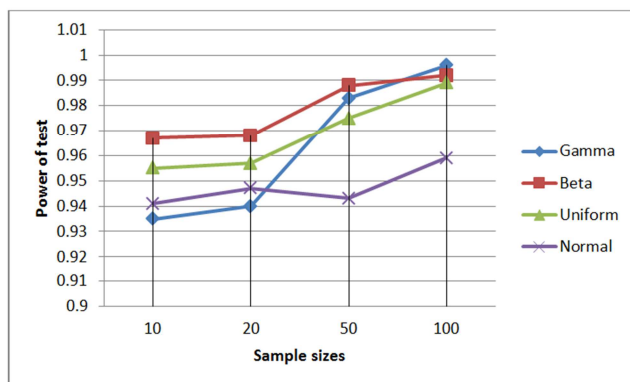Figure 8. Power of test of Lilliefores (LL) test.

Figure 9. *Power of test of Cramer–von Mises test.*

From figures 1 to 9, all the univariate normality test considered have significant power of test as the values are greater than 0.5 but less than 1.0. This implies researchers that formulated them have done a greater work for the formulation as all the methods show positive response in terms of application but the strength varies.

5. Conclusion

Test of normality plays a vital role in application of Statistics to data set as it prevents statistical error both type I and II errors. From the comparison, it is crystal clear that every test statistic for the determination of normality of a data set is adequate but the sensitivity vary significantly when sample size is taken into consideration. It was observed that shape of the data also affects the sensitivity of the technique. Care must be taken in considering any of the methods are uniformly most powerful but they are all locally most powerful.

6. Recommendation

Univariate normality tests must be used with great caution so as to prevent statistical error in the research.

Since there are other methods for the test of normality, it is therefore right to recommend the comparison of some other methods for their sensitivity.

Multivariate normality tests can also be compared to ease the strength of one by one test of normality in a large data set.

References

[1]   Abbas M. (2013): Robust Goodness of Fit Test Based on the Forward Search. American Journal of App. Mathematics and Statistics, 2013, Vol. 1, No. 1, 6-10.

[2]   Adefisoye J. O, Golam Kibria B. M, George F (2016) Performances of Several Univariate Tests of Normality: An Empirical Study. J Biom Biostat 7: 322. doi: 10.4172/2155-6180.1000322.

[3]   Agnieszka W, Robert D. I, Krzysztof B. (2020); Omnibus test for normality based on the Edgeworth expansion. PLoS ONE, vol. 15, no. 6, p. e0233901. Gale OneFile: Health and Medicine.

[4]   Anderson, T. W. (1962): On the Distribution of the Two-Sample Cramer–von Mises Criterion. The Annals of Mathematical Statistics (Institute of Mathematical Statistics) 33 (3): 1148–1159. doi: 10.1214/aoms/1177704477. ISSN 0003-4851.

[5]   Baghban A. A., Younespour S., Jambarsang S., Yousefi M., Zayeri F., and Jalilian F. A. (2013): How to test normality distribution for a variable: a real example and a simulationstudy. Journal of Paramedical Sciences (JPS). Vol. 4, No. 1 ISSN 2008-4978.

[6]   Bruno E. and Norbert H. (2020). Tests for multivariate normality -- a critical review with emphasis on weighted L^2-Statistics. arXiv; 2004.07332 Math.ST.

[7]   D'Agostino R. B. and Stephens M. A. (1986): Goodness-of-fit techniques. New York, Marcel Dekker.

[8]   Derya O., Atilla H. E., and Ersoz T. (2006): Investigation of Four Different Normality Tests in Terms of Type I Error Rate and Power Under Different Distributions. Turk. J. Med. Sci. 2006; 36 (3): 171-176.

[9]   Douglas G. B. and Edith S. (2002): A test of normality with high uniform power. Journal of Computational Statistics and Data Analysis 40 (2002) 435–445. www.elsevier.com/locate/csda.

[10]   Everitt, B. S. (2006): The Cambridge Dictionary of Statistics. Third Edition. Cambridge University Press, Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, Sao Paulo. Pp. 240-241.

[11]   Farrel P. J and Stewart K. R. (2006): Comprehensive Study of Tests for Normality and Symmetry; Extending the Spiegelhalter Test. Journal of Statistical Computation and Simulation. Vol. 76, No. 9, Pp 803-816.

[12]   Felix B. O and Senyo Y. A. (2016); Assessing Univariate and Multivariate Normality, A Guide For Non-Statisticians. Mathematical Theory and Modeling www.iiste.org ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online) Vol. 6, No. 2.

[13]   Frain J. C. (2006): Small Sample Power of Tests of Normality when the Alternative is an α –stable distribution. http://www.tcd.ie/Economics/staff/frainj/Stable_D istribution/normal.pdf

[14]   Guner B. and Johnson J. T. (2007): Comparison of the Shapiro-Wilk and Kurtosis Tests for the Detection of Pulsed Sinusoidal Radio Frequency Interference. The Ohio State University, Department of Electrical and Computer Engineering and Electro-Science Laboratory, 1320 Kinnear Road, Columbus, OH 43210, USA.

[15]   Gupta S. C. (2011): Fundamentals of Statistics. Sixth Revised and Enlarged Edition. Himilaya Publishing House PVT Ltd. Mumbai-400 004. Pp 16. 28-16.31.

[16]   Hadi Alizadeh Noughab (2018). A Comprehensive Study on Power of Tests for Normality. Journal of Statistical Theory and Applications, Vol. 17 (4), pp. 647–660. DOI: 10.2991/jsta.2018.17.4.7; ISSN 1538-7887. https://www.atlantis-press.com/journals/jsta

[17]   Jarque, C. M. and Bera, A. K. (1980): Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Economics Letters 6 (3): 255–259. doi: 10.1016/0165-1765(80)90024-5.

[18] Jarque, C. M. and Bera, A. K. (1981): Efficient tests for normality, homoscedasticity and serial independence of regression residuals: Monte Carlo evidence. Economics Letters 7 (4): 313–318. doi: 10.1016/0165-1765(81)900235-5.

[19] Jarque, C. M. and Bera, A. K. (1987): A test for normality of observations and regression residuals. International Statistical Review 55 (2): 163–172. JSTOR 1403192.

[20] Jason O. (2002): Notes on the use of data transformations. Practical Assessment, Research & Evaluation, 8 (6). http://PAREonline.net/getvn.asp?v=8&n=6.

[21] Jason O. (2010): Improving your data transformations: Applying the Box-Cox transformation. Practical Assessment, Research & Evaluation, 15 (12). Available online: http://pareonline.net/getvn.asp?v=15&n=12.

[22] Kuiper, N. H. (1960) Tests Concerning Random Points on a Circle. Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen, Series A, 63, 38-47. https://doi.org/10.1016/S1385-7258(60)50006-0

[23] Lilliefors HW (1967) On the Kolmogorov-Smirnov test for normality with mean and variance unknown. Journal of the American Statistical Association 62: 399-402.

[24] Lobato, Ignacio N. and Velasco, Carlos, (2004). "A Simple Test of Normality for Time Series". Econometric Theory, Cambridge University Press, vol. 20 (4), pages 671-689.

[25] Marmolejo-Ramos F. and Gonza´lez-Burgos J. (2012): A Power Comparison of Various Tests of Univariate Normality on Ex-Gaussian Distributions. European Journal of Research Methods for the Behavioural and Social Sciences. ISSN-Print 1614-1881 ISSN-Online 1614-2241. DOI: 10.1027/1614-2241/a000059. www.hogrefe.com/journals/methodology.

[26] Nasrin Khatun (2021). Applications of Normality Test in Statistical Analysis. Open Journal of Statistics 11 (01): 113-122. DOI: 10.4236/ojs.2021.111006.

[27] Nornadiah M. R. and Yap B. W (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. Journal of Statistical Modeling and Analytics. Vol. 2 No. I, 21-33.

[28] Ralph Henderson (2006). Testing experimental data for univariate normality Clinica Chimica Acta; International Journal of Clinical Chemistry 366 (1-2): 112-29. DOI: 10.1016/j.cca.2005.11.007.

[29] Shapiro SS, Wilk MB (1965) An analysis of variance test for normality. Biometrika 52: 591-611.

[30] Shapiro Wilk and Shapiro Francia (1972). An approximate analysis of variance test for normality. Journal of the American Statistical Association 67: 215-216.

[31] Shengyi & Mokhtarian, Patricia L & Johnston, Robert A., 2007. "Exploring the connections among job accessibility, employment, income, and auto ownership using structural equation modeling," Institute of Transportation Studies, Working Paper Series qt30v177dx, Institute of Transportation Studies, UC Davis.

[32] Stephen W. Looney (1995). How to Use Tests for Univariate Normality to Assess Multivariate Normality. The American Statistician, Volume 49, Issue 1. Pp 64-70.

[33] Thorsten Thadewald, and Herbert Büning, 2004. "Jarque-Bera test and its competitors for testing normality: A power comparison," Discussion Papers 2004/9, Free University Berlin, School of Business & Economics.

[34] Tanweer U. I. (2019); Ranking of Normality tests, An Appraisal through Skewed Alternative Space. Symmetry Basel. Volume 11 issue 7.

[35] Ukponmwan H. Nosakhare, and Ajibade F. Bright (2017). Evaluation of Techniques for Univariate Normality Test Using Monte Carlo Simulation. American Journal of Theoretical and Applied Statistics; 6 (5-1): 51-61http://www.sciencepublishinggroup.com/j/ajtasdoi: 10.11648/j.ajtas.s.2017060501.18ISSN: 2326-8999 (Print); ISSN: 2326-9006 (Online).

[36] Yap, B. W and Sim, C. H (2011). Comparisons of various types of normality tests. Journal of Statistical Computation and Simulation, 81: 12, 2141-2155, DOI: 10.1080/00949655.2010.520163.