

# Empirical Investigation of Type 1 Error Rate of Univariate Tests of Normality

Kayode Ayinde

Department of Statistics,  
Ladoke Akintola University of  
Technology,

P.M.B. 4000, Ogbomosho, Oyo State,  
Nigeria.

John Olatunde Kuranga

Department of Statistics  
Kwara State Polytechnic,  
P.M.B. 1375, Ilorin, Kwara State,  
Nigeria.

Gbenga Sunday Solomon

Department of Statistics,  
Ladoke Akintola University of  
Technology,  
P.M.B. 4000, Ogbomosho, Oyo State,  
Nigeria.

## ABSTRACT

Normality assumption is important in univariate parametric statistical tests. Either the variables or the error terms in the model have to be normally distributed before valid statistical conclusions could be made. Various tests of univariate normality including that of Pearson, Kolmogorov–Smirnov, Anderson-Darling, Shapiro–Wilk, Lilliefor, D’Agostino and Pearson, Jarque-Bera, Shapiro-Franca, Energy and Cramer-von Mises tests have been developed. However, when applied in practice, they hardly give the same result. Thus, this research work aims at investigating the Type 1 error rate of these tests so as to identify the best one and suggest the same for statistics users. The tests were compared by conducting Monte Carlo experiments five thousand (5000) times with six sample sizes at three pre-selected levels of significance. A test was adjudged good at a particular level of significance if its empirical Type 1 error rate approximated the true error rates most often. It is best if its number of counts at which it was good over the sample sizes and levels of significance was the highest. Results reveal that Type 1 error rate of all the univariate tests are good except that of Kolmogorov–Smirnov, Pearson Unadjusted and Jarque-Bera. Moreover, those of Anderson-Darling, Shapiro-Wilk, Energy and Cramer-von Mises tests are relatively best. They are therefore recommended for testing the assumption of normality in any univariate data set.

## Keywords

Parametric test statistics, Monte Carlo experiments, Type 1 error rate, Inferential statistics tests, Levels of significance.

## 1. INTRODUCTION

Normality assumption is an underlying assumption in statistical parametric tests. It is used in many statistical procedures including Regression analysis, Discriminant analysis and Analysis of Variance (ANOVA) and in virtually all the parametric statistical tests. Assessing the assumption of normality is required before proceeding with any relevant statistical inferences because it is the formulation upon which the interpretation of the results hold. There are three common techniques for checking the normality status of independent observations. These are graphical, numerical and the formal normality tests. The graphical is the easiest and it requires the normal quantile-quantile (Q-Qplot) and Histogram plots. Generally, graphical methods are informal approach. The Numerical methods include Skewness and Kurtosis indices which are generally referred to as standardized moments. The formal normality test is a scientific test in that tests methods are developed. The procedure involves testing whether a particular data set follows a normal distribution and computing the probabilities of how likely underlying data set is

normally distributed. In this study, attention is on univariate formal normality tests which are Unadjusted Pearson [1][2], Kolmogorov–Smirnov [3], Anderson-Darling [4], Shapiro–Wilk [5], Lilliefor [6], D’Agostino Skewness and Kurtosis [7], Adjusted Pearson [8], Omnibus [8], Jarque-Bera [9], Shapiro-Franca [10], Energy [11] and Cramer-von Mises [12] tests. This study focuses on evaluating their Type 1 error rates and comparatively distinguishes the best ones and suggests same for inferential usefulness.

## 2. REVIEW OF FORMAL UNIVARIATE TESTS OF NORMALITY

These are discussed under four major sub-headings as follows:

### a. Moment Tests

#### (1) Skewness and Kurtosis Tests

Let  $x_1, x_2, \dots, x_n$  be a random sample of  $n$  observations, then, the sample skewness and kurtosis,  $g_1$  and  $g_2$ , whose quantities consistently and respectively estimate the theoretical skewness and kurtosis of the distribution are defined as follows:

$$g_1 = \frac{m_3}{m_2^{3/2}} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left[ \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^{3/2}} \quad (1)$$

$$g_2 = \frac{m_4}{m_2^2} - 3 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left[ \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^2} - 3 \quad (2)$$

Where  $\bar{x}$  is the sample mean,  $m_2$ ,  $m_3$  and  $m_4$  are the second, third and fourth sample moments about the mean respectively. Pearson [2] derived the asymptotic distributions of skewness and kurtosis defined as follows:

$$Z_1 = \frac{g_1 - \mu_1(g_1)}{\sqrt{\mu_2(g_1)}} \sim N(0, 1) \quad (3)$$

$$Z_2 = \frac{g_2 - \mu_1(g_2)}{\sqrt{\mu_2(g_2)}} \sim N(0, 1) \quad (4)$$

Where  $\mu_1(g_1) = 0$ ,  $\mu_2(g_1) = \frac{6(n-2)}{(n+1)(n+3)}$ ,

$$\mu_1(g_2) = \frac{6}{n+1}, \mu_2(g_2) = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}$$

**(2) Omnibus  $K^2$  Test**

The Statistics  $Z_1$  and  $Z_2$  were combined by D'Agostino and Pearson [8] to produce an omnibus test defined as:

$$K^2 = Z_1(g_1)^2 + Z_2(g_2)^2 \sim \chi^2_2 \quad (5)$$

The test was developed to detect departure from normality due to either skewness or kurtosis [8].

**(3) Jarque-Bera Test**

The Jarque-Bera [9] test is a goodness-of-fit test of deviation from normality based on the sample skewness and kurtosis. The test is given as:

$$JB = \frac{n}{6} \left( g_1^2 + \frac{g_2^2}{4} \right) \sim \chi^2_2 \quad (6)$$

**B. Chi-Squared Test of Goodness of Fit**

Pearson [1] tested a null hypothesis stating that the frequency distribution of certain events observed in a sample is consistent with a particular theoretical distribution.

The test statistic is given as:

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{k-p} \quad (7)$$

Where  $O_i$  and  $E_i$  are the observed and expected frequency respectively,  $k$  is the number of cells in the table and  $p$  is the number of parameters estimated.

**C. Empirical Distribution Function (EDF) Tests**

**(1) Anderson-Darling Test**

The Anderson–Darling [4] proposed a test that is used to test whether or not a given sample of data is drawn from a given probability distribution. If the hypothesized distribution is  $F$  and empirical (sample) cumulative distribution function is  $F_n$ ; then, the quadratic empirical distribution function (EDF) statistics measure the distance between  $F$  and  $F_n$ . Anderson–Darling test is based on the distance given as:

$$EDF = n \int_{-\infty}^{\infty} (F_n(x) - F(x))^2 \left[ F(x)(1 - F(x)) \right]^{-1} dF(x) \quad (8)$$

$$A = n \int_{-\infty}^{\infty} \frac{(F_n(x) - F(x))^2}{F(x)(1 - F(x))} dF(x) \quad (9)$$

$$A^2 = -n - \left[ \sum_{i=1}^n \frac{2i-1}{n} [\ln(\Phi(Y_i)) + \ln(1 - \Phi(Y_{n+1-i}))] \right] \quad (10)$$

Where  $y_i$  is the  $i^{th}$  ordered observation from a distribution with cumulative distribution function CDF  $\Phi$

**(2) Cramer Von Mises Test**

The Cramér–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 [13]. It is defined as

$$\omega^2 = \int_{-\infty}^{\infty} [F_n(x) - F^*(x)]^2 dF^*(x) \quad (11)$$

Let  $x_1, x_2, \dots, x_n$  be the observed values, in increasing order. Then the statistic is given as:

$$T = n\omega^2 = \frac{1}{12n} + \sum_{i=1}^n \left[ \frac{2i-1}{2n} - F(x_i) \right]^2 \quad (12)$$

If this value is larger than the tabulated value, then the hypothesis that the data come from the distribution  $F$  can be rejected.

**(3) The Lilliefors Test**

Lilliefors [6] test is a normality test based on Kolmogorov test. It is a non-parametric test used to test the null hypothesis that data comes from a normal distributed population. The data consist of a random sample  $x_1, x_2, \dots, x_n$  size  $n$  associated with some unknown distribution, with unknown mean and unknown standard deviation.

The Lilliefors(LF) test is given as:

$$LF = \text{Sup}[F^*(x) - S(x)] \quad (13)$$

Where  $F^*(x)$  is the standard normal distribution function and  $S(x)$  is the empirical distribution function.

**D. Regression and Correlation Test**

Regression and Correlation tests are based on the ratio of two weighted least-square estimates of scale obtained from order statistics. The two estimates are the normally distributed weighted least squares estimates and the sample variance from other population.

**(1) Shapiro–Wilk test**

The Shapiro and Wilk [5] test is a tests for normality due to either skewness or kurtosis or both [14].

The test is:

$$W = \frac{\left( \sum_{i=1}^n a_i x_{(i)} \right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (14)$$

Where  $x_{(i)}$  is the  $i^{th}$  order statistic,  $\bar{x}$  is the sample mean, the constants  $a_i$  are given by [5] as:

$$a^T = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{1/2}} \quad (15)$$

Where  $m = E[x_{(1)}, x_{(2)}, \dots, x_{(n)}]$  and  $V = [\text{cov}(x_{(i)}, x_{(j)})]$  is the covariance matrix of those order statistics. Small value of  $W$  indicate normality.

**(2) Shapiro–Francia test**

Shapiro-Francia test [10] is simply the squared correlation between the ordered sample values and the (approximated) expected ordered quantiles from the standard normal distribution.

**(3) The D’Agostino D test**

D’Agostino [15] proposed the  $D$  test as an extension of the Shapiro–Wilk test. The D’Agostino proposed test eliminates the need to define the vector of weights  $a$  of the Shapiro–Wilk test and is obtained according to [16] as:

$$D = \sum_{i=1}^n \frac{(i - (n + 1)/2) \cdot x_{(i)}}{n^2 \cdot \sqrt{m_2}} \quad (16)$$

The normality hypothesis of the data is rejected for both small and large values of  $D$  using a two-sided test[16].

**3. METHODOLOGY**

In order to empirically investigate the Type 1 error rate of the normality tests, Monte Carlo experiments were conducted by generating data from normally distributed population five thousand times,  $X_i \sim N(0, 1), i = 1, 2, \dots, 5000$ , for six sample sizes namely;  $n = 10, 20, 30, 50, 100, 250$  and  $500$ . The R-Statistical software was used for the simulation study, and R-codes were written for all the thirteen (13) tests normality. The tests were compared at three (3) pre-selected levels of significance namely 0.1, 0.05 and 0.01. At a particular sample size, the number of times the true hypothesis is rejected is counted and the total is divided by the number of replication to estimate the Type 1 error rate of each test. The test whose error rate approximates the true error rate was considered good. That is test whose error rate fall into the preferred intervals as specified in Table 1 was considered good. Each

preferred intervals are such that the values therein approximate the true level of significance.

**Table 1: The True Level of Significance and Their Preferred Interval**

True levels of significance	Preferred interval
0.01	0.005 - 0.014
0.05	0.045 - 0.054
0.1	0.095 - 0.14

Source: [17] & [18]

Furthermore, the number of times each test was considered good was counted over the sample sizes and levels of significance. Thus, a total number of eighteen (18) counts was expected. A test was considered best if it has the highest number of total counts.

**4. RESULTS AND DISCUSSION**

**4.1 Results of Type 1 Error Rates of The Tests at 0.1 level of significance**

Table 2 shows the result of the Type 1 error rate at which each of the thirteen (13) tests of normality reject a true null hypothesis at 0.1 level of significance. In order to get a better view of their performance, the estimated Type 1 error rate closest to 0.1, using the preferred interval, are bolded, counted and presented. It is obvious from the results that all the tests generally have good Type 1 error rate except Kolmogorov–Smirnov, Pearson Unadjusted and Jarque–Bera. It should be noted that the Jarque-Bera test under estimated the Type 1 error rates. However, its error rate is not as bad as that of Pearson Unadjusted test. The Type 1 error rate of Kolmogorov–Smirnov test is relatively worst. Moreover, the Omnibus, skewness and kurtosis tests are also good except when the sample size is small,  $n=20$ . At all sample sizes, the Type 1 error rates of Anderson-Darling, Shapiro–Wilk, Energy and Cramer-Von Mises tests are relatively good (see Figure 1).

**Table 2: Type 1 Error Rate of The Tests at 0.1 Level of Significance**

Tests	Sample Size						Total Count
	20	30	50	100	250	500	
Anderson	<b>0.0986</b>	<b>0.102</b>	<b>0.1058</b>	<b>0.1036</b>	<b>0.1056</b>	<b>0.0982</b>	6
Kolmogorov	0.0002	0.0012	0.001	0.0006	0.0008	0.0006	0
Shapiro-Wilk	<b>0.0986</b>	<b>0.1016</b>	<b>0.1034</b>	<b>0.1036</b>	<b>0.1002</b>	<b>0.1054</b>	6
Cramer-vonMises	<b>0.0994</b>	<b>0.1048</b>	<b>0.1018</b>	<b>0.1012</b>	<b>0.1072</b>	<b>0.0972</b>	6
Shapiro-Franca	<b>0.099</b>	<b>0.1022</b>	<b>0.1018</b>	<b>0.1044</b>	<b>0.1016</b>	<b>0.1054</b>	6
Jarque-Bera	0.08	0.0908	0.081	0.0826	0.088	0.0886	0
Lilliefors	<b>0.0954</b>	<b>0.0972</b>	<b>0.1064</b>	<b>0.104</b>	<b>0.1004</b>	<b>0.1038</b>	6
Omnibus	0.0916	<b>0.0976</b>	<b>0.1002</b>	<b>0.0984</b>	<b>0.1018</b>	<b>0.1016</b>	5
Energy	<b>0.0984</b>	<b>0.1036</b>	<b>0.1032</b>	<b>0.1036</b>	<b>0.1072</b>	<b>0.1008</b>	6
Skewness	0.0934	<b>0.103</b>	<b>0.1014</b>	<b>0.1034</b>	<b>0.1076</b>	<b>0.1064</b>	5
Kurtosis	0.0928	<b>0.0952</b>	<b>0.0962</b>	<b>0.1008</b>	<b>0.1034</b>	<b>0.0976</b>	5
Pearson	<b>0.1218</b>	<b>0.122</b>	<b>0.1066</b>	<b>0.1052</b>	<b>0.1052</b>	<b>0.1002</b>	6
UnPearson	0.0336	0.0402	0.05	0.0558	0.057	0.065	0

Source: Simulation Results

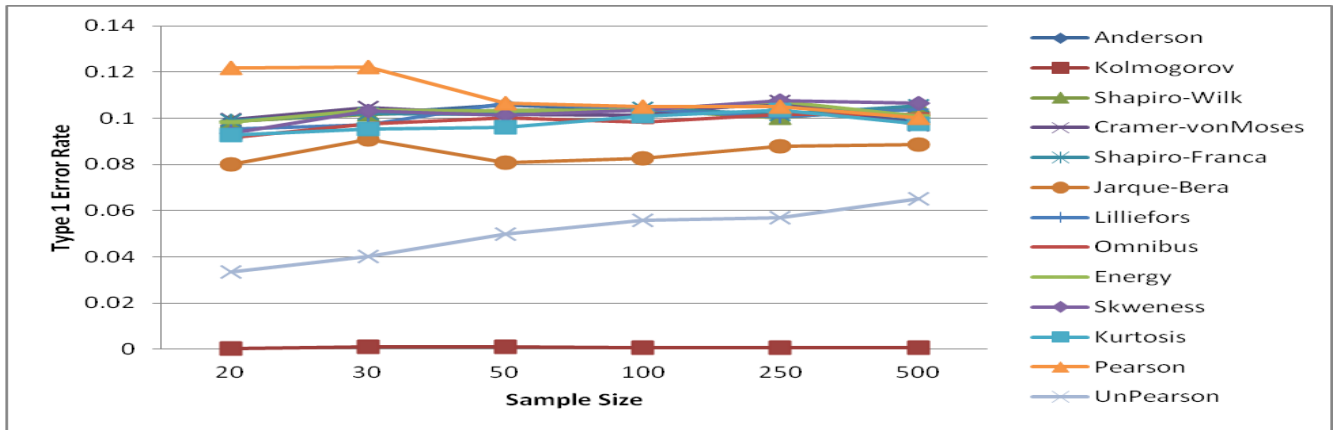


Figure 1: Type1 error rate of the Tests at 0.1 level of significance

### 4.2 Results of Type 1 Error Rates of the Tests at 0.05 Level of Significance

Table 3 displays the result of the Type 1 error rates of the tests of normality at 0.05 level of significance. Results show that all the tests generally have good Type 1 error rate except Kolmogorov–Smirnov, Pearson adjusted and Unadjusted, Omnibus and Jarque-Bera. Result further show that Jarque-

Bera test over estimated the Type 1 error rates except when the sample size is large,  $n \geq 250$ . Its error rate is also not as bad as that of Pearson’s Unadjusted test. The error rate of Skewness and Kurtosis tests are relatively better than that of Omnibus. Furthermore, the Type 1 error rate of Anderson-Darling, Shapiro–Wilk, Energy and Cramer-Von Mises tests are generally good at all the sample sizes (see Figure 2).

Table 3: Type 1 error rate of the Tests at 0.05 Level of Significance

Tests	Sample Size						Total Count
	20	30	50	100	250	500	
Anderson	<b>0.0476</b>	<b>0.0508</b>	<b>0.0516</b>	<b>0.0522</b>	<b>0.0498</b>	<b>0.0506</b>	<b>6</b>
Kolmogorov	0	0.0002	0	0	0	0	0
Shapiro-Wilk	<b>0.049</b>	<b>0.0502</b>	<b>0.0518</b>	<b>0.0504</b>	<b>0.0504</b>	<b>0.052</b>	<b>6</b>
Cramer-vonMises	<b>0.0488</b>	<b>0.0518</b>	<b>0.0516</b>	<b>0.051</b>	<b>0.05</b>	<b>0.0508</b>	<b>6</b>
Shapiro-Franca	<b>0.052</b>	<b>0.0546</b>	<b>0.0502</b>	<b>0.0522</b>	<b>0.0522</b>	0.0554	5
Jarque-Bera	0.0614	0.0668	0.058	0.0586	<b>0.0536</b>	<b>0.0478</b>	<b>2</b>
Lilliefors	<b>0.0488</b>	<b>0.0492</b>	<b>0.0546</b>	<b>0.0476</b>	0.0448	<b>0.0474</b>	<b>5</b>
Omnibus	0.0558	0.0592	<b>0.0544</b>	0.0556	<b>0.0534</b>	0.056	2
Energy	<b>0.0468</b>	<b>0.0498</b>	<b>0.052</b>	<b>0.052</b>	<b>0.051</b>	<b>0.0508</b>	<b>6</b>
Skewness	<b>0.0486</b>	<b>0.0508</b>	<b>0.0482</b>	<b>0.0496</b>	<b>0.054</b>	0.0574	5
Kurtosis	<b>0.0464</b>	<b>0.0478</b>	<b>0.0518</b>	0.057	<b>0.0488</b>	<b>0.0516</b>	<b>5</b>
Pearson	0.044	<b>0.0524</b>	0.056	0.0558	<b>0.0518</b>	0.0554	2
UnPearson	0.0136	0.0178	0.022	0.0252	0.0234	0.0286	0

Source: Simulation Results

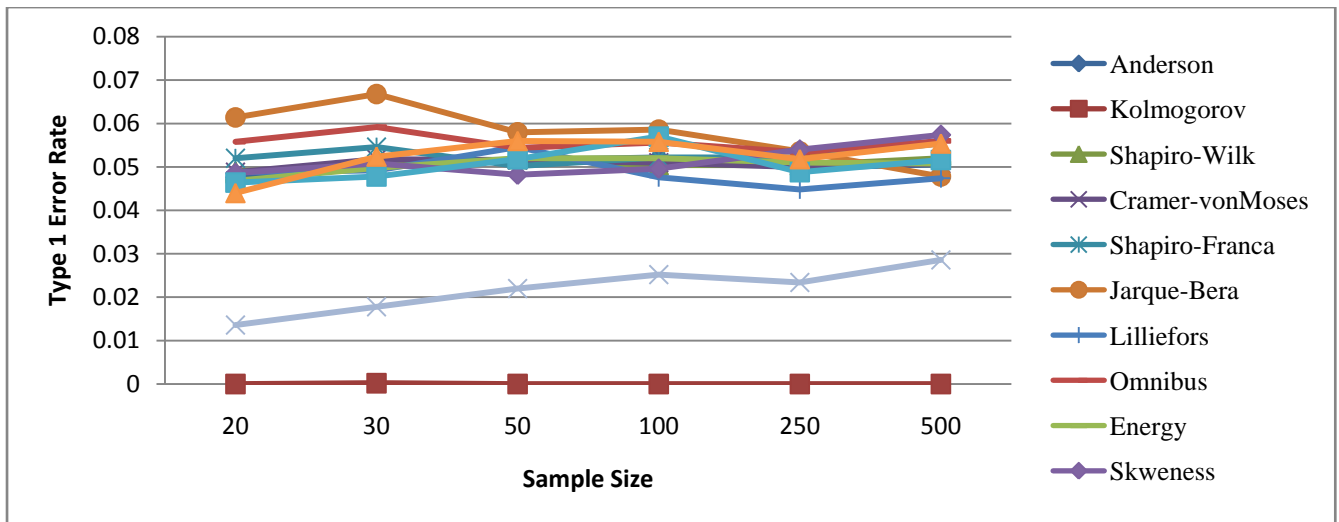


Figure 2: Type1 error rate of the Tests at 0.05 Level of Significance

### 4.3 Results of Type 1 Error Rates of the Statistics at 0.01 level of significance

A summary of the results of the empirical Type 1 error rate at 0.01 level of significance is presented in Table 4. In order to get a clearer understanding of the performance of the tests of normality at this level of significance, the value of the test whose Type 1 error rate is closest to 0.01, as a result of the

preferred interval, are bolded and counted. The results revealed that the tests as a whole have good Type 1 error rate except Kolmogorov–Smirnov, Pearson Unadjusted Omnibus and Jarque-Bera. Additionally, Jarque-Bera and Omnibus tests over estimate the Type 1 error rates while the Pearson Unadjusted test under estimate the error rate. However, Kolmogorov–Smirnov test is comparatively not good. This is further shown in Figure 3.

Table 4: Type 1 Error Rate of The Tests at 0.01 Level of Significance

Tests	Sample Size						Total Count
	20	30	50	100	250	500	
Anderson	<b>0.0084</b>	<b>0.0094</b>	<b>0.0114</b>	<b>0.0104</b>	<b>0.009</b>	<b>0.0108</b>	<b>6</b>
Kolmogorov	0	0	0	0	0	0	0
Shapiro-Wilk	<b>0.0096</b>	<b>0.01</b>	<b>0.0084</b>	<b>0.01</b>	<b>0.0114</b>	<b>0.012</b>	<b>6</b>
Cramer-vonMises	<b>0.0088</b>	<b>0.0098</b>	<b>0.0108</b>	<b>0.0088</b>	<b>0.0088</b>	<b>0.0108</b>	<b>6</b>
Shapiro-Franca	<b>0.0106</b>	<b>0.0118</b>	<b>0.009</b>	<b>0.0122</b>	<b>0.0126</b>	<b>0.0124</b>	<b>6</b>
Jarque-Bera	0.0364	0.04	0.0304	0.0288	0.0232	0.0162	0
Lilliefors	<b>0.01</b>	<b>0.01</b>	<b>0.0128</b>	<b>0.01</b>	<b>0.0106</b>	<b>0.0088</b>	<b>6</b>
Omnibus	0.0202	0.019	0.0154	0.018	0.0178	0.0138	0
Energy	<b>0.009</b>	<b>0.0098</b>	<b>0.012</b>	<b>0.0108</b>	<b>0.0096</b>	<b>0.011</b>	<b>6</b>
Skewness	<b>0.0104</b>	<b>0.0106</b>	<b>0.008</b>	<b>0.0108</b>	<b>0.0122</b>	<b>0.01</b>	<b>6</b>
Kurtosis	<b>0.0084</b>	<b>0.0112</b>	<b>0.0122</b>	0.0152	<b>0.0132</b>	<b>0.0118</b>	<b>5</b>
Pearson	<b>0.0098</b>	<b>0.0134</b>	<b>0.011</b>	<b>0.011</b>	<b>0.0086</b>	<b>0.009</b>	<b>6</b>
UnPearson	0.0024	0.0034	0.004	0.0034	0.0034	0.004	0

Source: Simulation Results

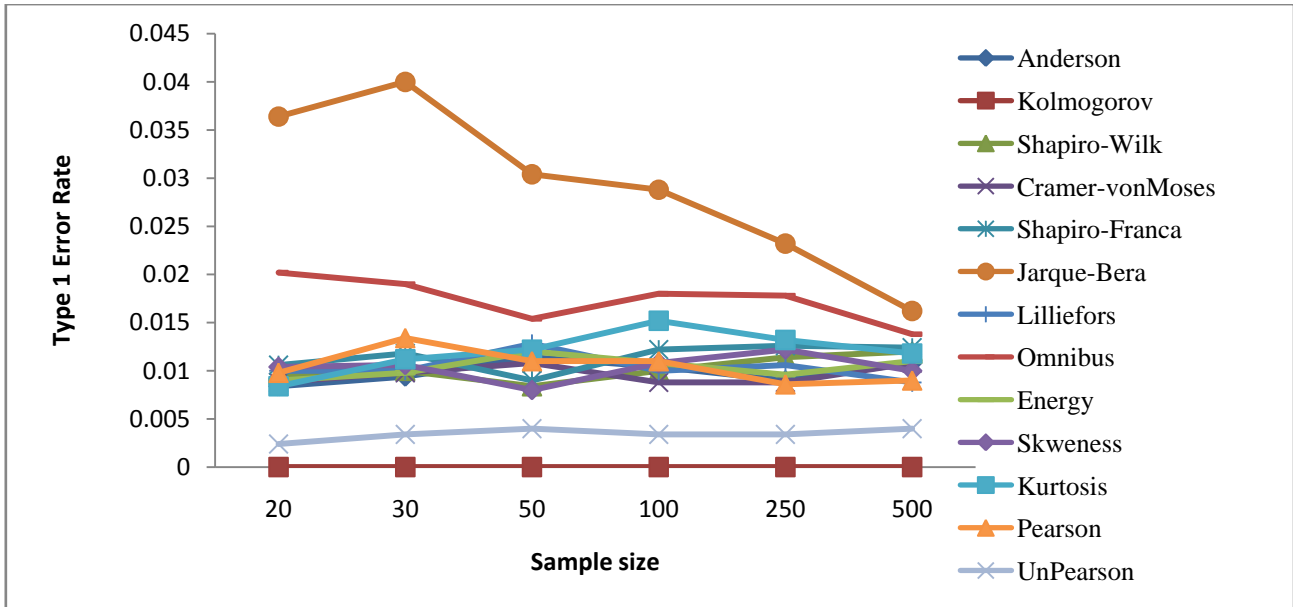


Figure 3: Type 1 Error Rate of the Tests at 0.01 Level of Significance

#### 4.4 Overall performance of the Normality Test Statistics

In order to see the performance of the statistics clearly, the number of times the Type 1 error rate falls into the preferred interval (good performance) is counted over the three levels of significance and the sample sizes. This is summarized in Table 5. Generally, the Type 1 error rate of Anderson-Darling, Shapiro-Wilk, Energy, Cramer-von Mises are best since they

have the highest number of times the estimated Type 1 error rate fall into the preferred intervals. These are followed by those of Shapiro-Franca and Lilliefors tests (see Figure 4). Consequently, Anderson-Darling, Shapiro-Wilk, Energy and Cramer-Von Mises tests are recommended for use in test of normality of a data set.

Table 5: Number of good performance of the Normality Test

Tests	Sample Size						Total counts
	20	30	50	100	250	500	
Anderson	3	3	3	3	3	3	18
Kolmogorov	0	0	0	0	0	0	0
Shapiro-Wilk	3	3	3	3	3	3	18
Cramer-von Mises	3	3	3	3	3	3	18
Shapiro-Franca	3	3	3	3	3	2	17
Jarque-Bera	0	0	0	0	1	1	2
Lilliefors	3	3	3	3	2	3	17
Omnibus	0	1	2	1	2	1	7
Energy	3	3	3	3	3	3	18
Skweness	2	3	3	3	3	2	16
Kurtosis	2	3	3	1	3	3	15
Pearson	2	3	2	2	2	3	14
UnPearson	0	0	0	0	0	0	0

Source: Table 2, 3 and 4.

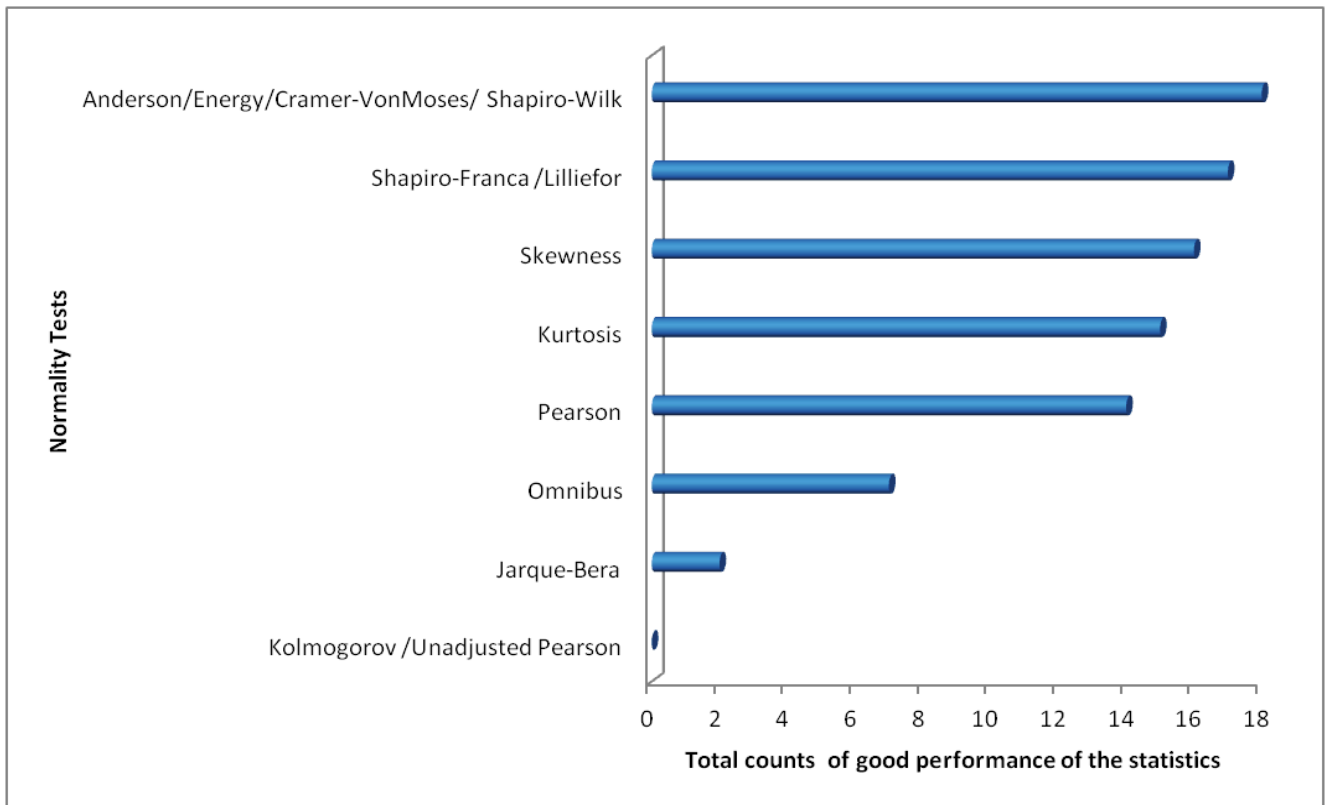


Figure 4: Bar chart showing Total Number of good performance of the Normality Test Statistics

## 5. CONCLUSION

In this study, the Type 1 error rates of the univariate tests of normality have been empirically examined. Results have revealed that Type 1 error rate of Anderson-Darling, Shapiro-Wilk, Energy, Cramer-von Mises tests are relatively best. These are followed by those of Shapiro-Franca and Lilliefors tests. But that of Skewness, Kurtosis, Pearson and Omnibus tests, in this order are fairly good. The performance of Jarque-Bera test is comparatively not good while that of Unadjusted Pearson and Kolmogorov-Smirnov are the least. Findings have shown that graphical methods (Q-Q plot, Histogram, Stem and Leaf plot and box plot) usually provide information about the shape of the distribution of the data but do not assure that the set of data is normal. Consequently, Anderson-Darling, Shapiro-Wilk, Energy and Cramer-Von Mises formal tests of normality are therefore recommended for practitioners.

## 6. REFERENCES

- [1] Pearson, K. 1900. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, *Philosophical Magazine Series*, 550 (302), 157-175.
- [2] Pearson, K. 1905. On general theory of skew correlation and Non-linear regression, London: Dulau and Co.
- [3] Kolmogorov, A. N. 1933. Sulla determinazione empirica di una legge di distribuzione, *Giornale dell Istituto Italiano degli Attuari*, 4, 83-91.
- [4] Anderson, T. W. and Darling, D. A. 1954. A Test of Goodness of Fit, *Journal of the American statistical Association*, 49 (268), 765-769.
- [5] Shapiro, S. S. and Wilk, M. B. 1965. An Analysis of variance Test for Normality, *Biometrika*, 52 (3), 591 – 611.
- [6] Lilliefors, H. W. 1967. On the Kolmogorov – Smirnov Test for Normality with mean and variance unknown, *Journal of American statistical Association*, 62 (318), 399-402.
- [7] D’Agostino, 1970. R. B. Transformation to normality of the null distribution of  $g_1$ , *Biometrika*, 57, 679-681.
- [8] Agostino, R. D and Pearson, E. S. 1973. Test for Departure from Normality. Empirical Results for the Distributions of  $b_2$  and  $\sqrt{b_1}$ , *Biometrika*, 60 (3), 613-622.
- [9] Jarque, C. M. and Bera, A. K. 1987. A test for Normality of observations and regression residual, *Internat. Statist.*, 55 (2), 163 – 172.
- [10] Shapiro, S. S. and Francia, R. S. 1972. An Approximate analysis of variance test for normality, *Journal of American Statistical Association*, 67, 215-216.
- [11] Szekely, G. J. and Rizzo, M. L. 2005. A New Test for Multivariate Normality, *Journal of Multivariate Analysis*, 93 (1), 58-80.
- [12] Thadewald, T. and Buning, H. 2007. Jarque – Bera and its Competitors for Testing Normality, *Journal of Applied Statistics*, 34 (1), 87 – 91.
- [13] Anderson, T. W. 1962. On the Distribution of the Two-Sample Cramer-von Mises Criterion, *The Annals of Mathematical Statistics (Institute of Mathematical Statistics)* doi:10.1214/aoms/1177704477. ISSN 0003-4851, 33 (3), 1148-1159.

- [14] Althous, I. A., Ware W. B. and Ferron, J. M. 1998. Detecting departures from Normality; A Monte carlo on simulation of A New omnibus Test base Moments, in Annual Meeting of the American Education Research Association, San Diego, CA.
- [15] D'Agostino, R. B. 1971. An omnibus test of normality for moderate and large size samples, *Biometrika*, 52 ( 2), 341–348.
- [16] Romao, X., Delgado, R. and Costa, A. 2010. An Empirical Power Comparison of Univariate Goodness-of-Fit Test for Normality, *Journal of Statistical Computation and Simulation*, 80 (5), 545-591.
- [17] Kuranga, J. O. 2015. Type 1 Error Rate and Power Comparison of Some Normality Test Statistics, Ogbomosho: Unpublished M.Phil. Thesis, Department of Statistics, Ladoke Akintola University of Technology.
- [18] Ayinde, K., Solomon, G. S. and Adejumo, A. O. 2016. A Comparative Study of the Type 1 Error Rates of Some Multivariate Tests of Normality, *Nigerian Association of Mathematical Physics*, 34 (1), 185-198.