Alabama Journal of Mathematics

Alabama Journal of Mathematics
Volume 45 (1) (2022), 25 – 36
DOI : 10.10004/ajm

Original Article

Univariate Normality Tests Based on Skewness and Kurtosis: A Monte Carlo Simulation

Danush K. Wijekularathna[1], N. A. Newman[1], Sukhitha W. Vidurupola[2]

[1] *Department of Mathematics and Statistics,*
Troy University
Troy, AL 36082, USA
[2] *Rogers State University*

Abstract

A comparison of eight different univariate normality tests, including interesting and most recently developed tests, was carried out using Monte Carlo simulations for distributions with different skewness and kurtosis under four groups: symmetric long-tailed, symmetric short-tailed, asymmetric, and normal modified distributions. Several sample sizes were used. Results suggest what univariate normality tests under consideration are best, based on the nature of the nonnormality and sample size.

Mathematics Subject Classification (2020). 57K10, 57K18

Keywords. Univariate normality, 2nd-power skewness, 2nd-power kurtosis, Upper 5 percentiles, Power comparison

1. Introduction

Assessing univariate normality is an important task in the analysis of univariate data. This is due to the fact that most parametric statistical techniques have been developed based on normality assumption. Thus, testing the normality assumption is an essential part of the univariate data analysis.

There are two main types of methods that can be used to test normality. One is to use graphical methods in which the researcher observes graphs and charts such as histograms, box plots, and Q-Q plots to evaluate the normality assumption. However, these graphical methods may be appropriate only if the statistical procedure used is robust to the normality assumption. When the given statistical procedure is not robust, more rigorous statistical testing methods should be used to evaluate the normality assumption.

In the literature of univariate analysis, there are many different tests devoted to this problem. Some references are [1], [18], [12], [6], [17], [16] , [11], [5], [20], [15] and many others.

Email addresses: dwijekularathna@troy.edu,
newman@troy.edu
Received: August 10, 2020; Accepted: January 11, 2021

There are several studies that have been carried out in the literature addressing the comparison of these univariate normality tests. [2] compared ten univariate normality tests that were used from three different categories: empirical distribution function tests, regression and correlation tests, and omnibus skewness and kurtosis tests. The power of these tests were computed using Monte Carlo samples of small, moderate, and large sample sizes from symmetric, skewed, contaminated and mixed distributions. Based on his simulation results, Seier claimed that the Shapiro-Wilk and the Chen-Shapiro tests have higher average power, and most crucial and widely known Empirical distribution tests are Kolmogovov-Smirnov, Anderson-Darling, and Cramer-von Mises tests.

[14] compared the power of four formal tests of normality: Shapiro-Wilk test, Kolmogovov-Smirnov test, Lilliefors test, and Anderson-Darling test. They generated Monte Carlo simulations of sample data from alternative distributions that follow symmetric and asymmetric distributions. They concluded that the Shapiro-Wilk test is the best test to be adopted for both symmetric non-normal and asymmetric distributions.

[21] compared the power of eight selected normality tests: Shapiro-Wilk test, Kolmogovov-Smirnov test, Lilliefors test, Cramer-von Mises test, Anderson-Darling test, D'Agostino-Pearson test, Jarque-Bera test, and chi-squared test. They considered alternative distributions that follow symmetric short-tailed, Symmetric long-tailed, and asymmetric distributions in Monte Carlo Studies. [21] provide a brief description of the eight test statistics and suggest that the D'Agostino and Shapiro-Wilk tests perform better for symmetric short-tailed distributions, the Jarque-Bera and D'Agostino tests perform better for symmetric long-tailed distributions, and the Shapiro-Wilk test is the most powerful test for asymmetric distributions.

[13] compared seven normality tests: Kolmogovov-Smirnov, Anderson-Darling, Kuiper, Jarque-Bera, Cramer-von Mises, Shapiro-Wilk, and Vasicek. They computed empirical power for each test by Monte Carlo simulation methods under twenty alternatives, which were divided into four subgroups, depending on the support and shape of their densities. In their conclusions, out of these seven tests, they recommend to use the Jarque-Bera test statistic for symmetric distributions and Shapiro-Wilk test for asymmetric distributions, in practice.

[20] comprehensively discussed and summarized a collection of normality tests, which were compared based on power and ease of use. In addition, he discussed the performance of the tests in the presence of outliers as well.

In this study, we used currently-available, including recently developed, eight univariate tests and compared them to investigate what univariate normality tests are the best depending on the nature of the non-normality and sample size. Monte Carlo simulations were used to investigate the strengths and weaknesses of the considered test statistics. To calculate the empirical power of the tests, alternative distributions were used from four different categories: symmetric long-tailed, symmetric short-tailed, asymmetric, and normal modified distributions. Section 2 briefly discusses each of the eight tests being compared and Section 3 presents power comparisons of the tests using Monte Carlo simulations for alternative distributions for several sample sizes. Finally, Section 4 provides conclusions and recommendations from the investigation.

2. Description of test statistics

We investigate eight normality tests based on univariate skewness which is a measure of symmetry about the mean for a probability density, and kurtosis which is regarded as a measure of the peakedness of a probability density for a random variable. Skewness equal to zero indicates that the probability density is perfectly symmetric about its mean. Kurtosis is intended to measure the height and sharpness of a dominant central peak, relative to that of a standard bell-shaped curve. Skewness and Kurtosis of a standard

Normal distribution are 0 and 3.0 respectively. In this section, we briefly describe the eight tests under consideration: Z_{ED} test, X_{AD} test, Gel-Gastwirth test, Brys-Hubert-Struyf test, Bonett-Seier test, Doornik-Hansen test, Jaque-Bera test, and D'Agostino-Pearson test.

2.1. Z_{ED} and X_{AD} Tests of Normality

[7] defined the 2nd-power skewness and kurtosis, which are interesting alternatives to the classical Pearson's skewness and kurtosis, as follows.

For a sample $X_1, X_2, ..., X_n$, '2nd-power skewness' and '2nd-power kurtosis' are denoted by B_2 and K_2 and defined as

$$B_2 = \frac{1}{n}\sum_{i=1}^{n} z_i^2 sign(Z_i) \tag{2.1}$$

$$K_2 = \frac{1}{n}\sum_{i=1}^{n} z_i^2 log|Z_i| \tag{2.2}$$

where

$$Z_i = S_n^{-1}(X_i - \bar{X}_n), S_n = \Big[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2\Big]^{1/2}, \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Sample 2nd-power skewness and kurtosis are used to build an univariate test of normality that can also be derived as Rao's score test on the asymmetric power distribution, which combines the large range of exponential tail behavior provided by the exponential power distribution family with various levels of asymmetry.

The 'transformed 2nd-power skewness' is denoted by $Z(B_2)$ and defined as

$$Z(B_2) = \frac{n^{1/2}B_2}{[(3-8/\pi)(1-1.9/n)]^{1/2}}, \tag{2.3}$$

where B_2 is given in equation (2.1) .

The 'transformed 2nd-power net kurtosis' is denoted by $Z(K_2 - B_2^2)$ and defined as

$$Z(K_2 - B_2^2) = \frac{n^{1/2}[(K_2 - B_2^2)^{1/3} - ((2 - log2 - \gamma)/2)^{1/3}(1 - 1.026/n)]}{[72^{-1}((2 - log2 - \gamma)/2)^{-4/3}(3\pi^2 - 28)(1 - 2.25/n^{0.8})]^{1/2}}, \tag{2.4}$$

where $\gamma = 0.577215665...$ and B_2 and K_2 are given in equations (2.1) and (2.2), respectively.

Finally, the 'transformed 2nd-power kurtosis' is denoted by $Z(K_2)$ and defined as

$$Z(K_2) = \frac{n^{1/2}[((2K_2)^{\alpha_n} - 1)/\alpha_n + ((2 - log2 - \gamma)^{-0.06} - 1)/0.06 + 1.32/n^{0.95}]}{[(2 - log2 - \gamma)^{-2.12}(3\pi^2 - 28)/2 - 3.78/n^{0.733}]^{1/2}} \tag{2.5}$$

where K_2 is given in equation (2.2), $\alpha_n = -0.06 + 2.1/n^{n^{0.67}}$.

The two univariate statistics proposed by [7] can be presented as follows.

The first univariate test statistic is denoted by X_{AD} and given by

$$X_{AD} = Z^2(B_2) + Z^2(K_2 - B_2^2) \tag{2.6}$$

where the transformed 2nd-power skewness $Z(B_2)$ and the transformed 2nd-power net kurtosis $Z(K_2 - B_2^2)$ are given in the equations (2.3) and (2.4), respectively. Furthermore, under the null hypothesis, X_{AD} has approximately a χ_2^2 distribution for all $n \geq 10$.

The second statistic they proposed to test the composite hypothesis of normality, for finite sample sizes $n \geq 10$, is denoted by Z_{ED} and given by

$$Z_{ED} = Z(K_2) \tag{2.7}$$

where the transformed 2nd-power kurtosis $Z(K_2)$ is given in equation (2.5). Furthermore, under the null hypothesis, Z_{ED} is approximately a standard normal distribution for all $n \geq 10$.

2.2. The Gel-Gastwirth Test for Normality

[9] utilized a robust estimate of spread which is less influenced by outliers and proposed the new Jarque-Bera (RJB) test statistic:

$$RJB = \frac{n}{6}\Big(\frac{m_3}{J_n^3}\Big)^2 + \frac{n}{64}\Big(\frac{m_4}{J_n^4} - 3\Big)^2 \tag{2.8}$$

where

$$J_n = \sqrt{\frac{\pi}{2n^2}}\sum_{i=1}^{n}|X_i - M|, \; m_j = n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^j \text{ and } M \text{ is the sample median.}$$

2.3. The Brys-Hubert-Struyf Test for Normality

[4] proposed a goodness-of-fit test based on robust measures of skewness and tail weight. They considered the medcouple (MC), a robust skewness measure, proposed in Brys et al.[10], [3] and defined as:

$$MC(F) = med_{X_{(i)} \leq m_F \leq X_{(j)}} h(x_{(i)}, x_{(j)}) \tag{2.9}$$

with $x_{(i)}$ and $x_{(j)}$ sampled from F, $m_F = F^{-1}(0.5)$, and the kernel function h given by

$$h(x_{(i)}, X_j) = \frac{(x_{(j)} - m_F) - (m_F - x_{(i)})}{(x_{(j)} - x_{(i)})}.$$

Furthermore, they considered the left medcouple (LMC) and right medcouple (RMC) of the left and right tail weight measures and defined as follows:

$$LMC(F) = -MC(x < m_F)$$

and

$$RMC(F) = MC(x > m_F).$$

The Brys-Hubert-Struyf test statistic is defined by

$$T_{ML} = n(\omega - w)^t \Sigma^{-1}(\omega - w) \tag{2.10}$$

where $w = [MC, LMC, RMC]^t = [0, 0.199, 0.199]^t$ and

$$\Sigma = \begin{bmatrix} 1.2500 & 0.3230 & -0.3230 \\ 0.3230 & 2.6200 & -0.0123 \\ -0.3230 & -0.0123 & 2.6200 \end{bmatrix}.$$

T_{ML} test statistic approximately follows a chi-square distribution with three degrees of freedom.

2.4. Bonett-Seier Test for Normality

[2] proposed a modified measure of kurtosis for testing normality assumption. The Bonett and Seier normality test statistic is defined as:

$$T_w = \frac{\sqrt{n-2}(\hat{\omega}-3)}{3.54} \tag{2.11}$$

where

$$\hat{\omega} = 13.29[ln\sqrt{m_2} - ln(n^{-1}\sum_{i=1}^{n}|X_i - \bar{X}|)]$$

T_w approximately follows a standard normal distribution.

2.5. Doornik-Hansen Test for Normality

[8] suggested an easy to use version of the omnibus test for normality using skewness and kurtosis based on [19]. Let Z_1 and Z_2 denote the transformed skewness and kurtosis. The test statistic is then defined as

$$E_p = Z_1^2 + Z_2^2 \tag{2.12}$$

where

$Z_1 = \delta_1 \; log\{y + (y^2+1)^{1/2}\}$, $Z_2 = \left\{\left(\frac{\chi}{2\alpha}\right)^{1/3} - 1 + \frac{1}{9\alpha}\right\}(9\alpha)^{1/2}$, $\delta_1 = \frac{1}{\{log \;\; \omega\}^{1/2}}$,
$y = \sqrt{b_1}\left\{\frac{\omega^2-1}{2}\frac{(n+1)(n+3)}{6(n-2)}\right\}^{1/2}$, $\chi = (b_2-1-b_1)2k$, $\alpha = a+b_1c$, $\omega^2 = -1+\{2(\beta-1)\}^{1/2}$,

$$b_1 = \frac{m_3}{m_2^{3/2}}, \tag{2.13}$$

$$b_2 = \frac{m_4}{m_2^2}, \tag{2.14}$$

$$a = \frac{(n-2)(n+5)(n+7)(n^2+27n-70)}{6\delta_2},$$

$$c = \frac{(n-7)(n+5)(n+7)(n^2+2n-5)}{6\delta_2},$$

$$k = \frac{(n+5)(n+7)(n^3+37n^2+11n-313)}{12\delta_2},$$

$$\beta = \frac{3(n^2+27n-70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)},$$

$$\delta_2 = (n-3)(n+1)(n^2+15n-4),$$

and

$$m_j = n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^j. \tag{2.15}$$

This test is approximately chi-squared distributed with two degrees of freedom.

2.6. Jaque-Bera Test for Normality

[11] proposed a normality test that is simple in computations and is asymptotically efficient. Jarque-Bera test statistic can be defined as

$$JB = \frac{n}{6}\Big(b_1 + \frac{(b_2-3)^2}{4}\Big)$$

where b_1 and b_2 are same as in the equations (2.13) and (2.14), respectively. JB test statistic asymptotically follows a chi-squared distributed with two degrees of freedom.

2.7. D'Agostino-Pearson Test for Normality

[6] proposed a normality test that is based on skewness and kurtosis. Their proposed test statistic is given by

$$OT = TS^2 + TK^2.$$

TS and TK are given by

$$TS = \delta\ \ln(Y/\alpha + \Big\{(Y/\alpha)^2 + 1\Big\}^{1/2}) \tag{2.16}$$

where

$$Y = \sqrt{b_1}\left\{\frac{(n+1)(n+3)}{6(n-2)}\right\}^{1/2},$$

$$\alpha = \left\{\frac{2}{(W^2-1)}\right\}^{1/2},$$

$$\delta = \frac{1}{\sqrt{\ln\ W}},$$

$$W^2 = \{2(\beta_2-1)\}^{1/2} - 1,$$

$$\beta_2 = \frac{3(n^2+27n-70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)},$$

and

$$TK = \frac{(1-\frac{2}{9A}) - [\frac{1-2/A}{1+x\sqrt{2/(A-4)}}]^{1/3}}{\sqrt{2/(9A)}}$$

where

$$x = \frac{b_2 - 3(n-1)/(n+1)}{24n(n-2)(n-3)/[(n+1)^2(n+3)(n+5)]},$$

$$A = 6 + \frac{8}{\sqrt{\beta_1}}\left[\sqrt{1+\frac{4}{\sqrt{\beta_1}}} + \frac{2}{\sqrt{\beta_1}}\right],$$

and

$$\sqrt{\beta_1} = \frac{6(n^2-5n+2)}{(n+7)(n+9)}\sqrt{\frac{6(n+3)(n+5)}{n(n-2)(n-3)}}.$$

OT statistic has approximately a chi-squared distribution with two degree of freedom. Please note that b_1 and b_2 are same as in the equations (2.13) and (2.14), respectively.

3. Simulation Study

In this paper eight different tests of univariate normality were comprehensively compared for accuracy, upper percentiles, type I error, and power. The tests compared were Z_{ED}, X_{AD}, Gel-Gastwirth, Brys-Hubert-Struyf, Bonett-Seier, Doornik-Hansen, Jaque-Bera, and D'Agostino-Pearson test. Each test was compared under various alternatives: four groups of distributions (symmetric long-tailed, symmetric short-tailed, asymmetric, normal modified distributions), different significance levels ($\alpha = 0.05, 0.10$), and different sample sizes ($n = 10, 15, 20, 30, 50, 75, 100, 200, 300, 500$) using Monte Carlo simulations with $10,000$ samples from each sample size.

3.1. Upper Percentiles

2*n	Test Statistics							
	X_{AD}	Z_{ED}	RJB	T_{ML}	T_w	E_p	JB	OT
10	4.841 (6.214)	1.302 (1.705)	3.320 (7.008)	3.347 (4.200)	0.977 (1.462)	4.715 (6.007)	1.660 (2.642)	4.507 (6.729)
20	4.562 (5.939)	1.255 (1.604)	3.672 (6.991)	4.112 (5.120)	1.100 (1.558)	4.452 (5.835)	2.328 (3.728)	4.442 (6.237)
30	4.543 (5.924)	1.250 (1.628)	3.953 (7.262)	5.624 (7.007)	1.131 (1.591)	4.401 (5.848)	2.813 (4.286)	4.581 (6.205)
50	4.545 (5.921)	1.248 (1.612)	3.938 (6.781)	5.863 (7.477)	1.169 (1.569)	4.417 (5.858)	3.134 (4.857)	4.491 (6.219)
75	4.575 (5.921)	1.256 (1.598)	3.921 (6.467)	5.775 (7.180)	1.190 (1.611)	4.377 (5.803)	3.461 (5.193)	4.567 (6.163)
100	4.524 (5.985)	1.251 (1.637)	4.043 (6.398)	5.875 (7.384)	1.201 (1.618)	4.383 (5.967)	3.672 (5.385)	4.641 (6.106)
200	4.553 (5.934)	1.271 (1.617)	4.010 (5.831)	6.063 (7.549)	1.221 (1.624)	4.346 (5.744)	3.950 (5.455)	4.508 (5.994)
300	4.568 (5.930)	1.288 (1.628)	4.056 (5.761)	6.066 (7.540)	1.242 (1.620)	5.710 (4.442)	4.093 (5.464)	4.533 (5.962)
500	4.582 (5.884)	1.275 (1.639)	4.186 (5.624)	6.165 (7.694)	1.246 (1.595)	4.423 (5.818)	4.220 (5.573)	4.494 (5.7880
	4.605(5.991)	1.281(1.644)	4.605(5.991)	6.251(7.814)	1.281(1.644)	4.605(5.991)	4.605(5.991)	4.605(5.9915)

Table 1. Upper 5(10) Percentiles

The values of upper 5 and 10 percentiles of the test statistics under consideration are given in Table 1. The lowest row of each table shows theoretical (asymptotic) values for the corresponding distribution of test statistics. When sample size increases, values of all statistics seem to converge to their asymptotic distribution values. The statistics X_{AD} and Z_{ED} start to converge to their asymptotic distribution values faster than others, when the sample size is about 20. Statistic OT starts to converge when the sample size is about 75. When the sample size is small (n=10, 20, 30), JB values are smaller compared to its asymptotic distribution values. Overall, it seems, X_{AD} has the best accuracy for each sample size and Z_{ED} has the second best.

In addition, for both upper 10 and 5 percentiles, OT has a better accuracy than T_w. The statistic T_w is closer to its asymptotic distribution values than T_{ML}, T_{ML} has a better accuracy than E_P, E_p has a better accuracy than RJB, and RJB has a better accuracy than JB.

3.2. Type I Errors

2*n	Test Statistics							
	X_{AD}	Z_{ED}	RJB	T_{ML}	T_w	E_p	JB	OT
10	0.050 (0.100)	0.051 (0.100)	0.050 (0.101)	0.050 (0.100)	0.051 (0.101)	0.051 (0.100)	0.051 (0.102)	0.051 (0.100)
15	0.050 (0.099)	0.051 (0.099)	0.050 (0.099)	0.051 (0.101)	0.049 (0.099)	0.051 (0.100)	0.050 (0.101)	0.050 (0.100)
20	0.051 (0.101)	0.050 (0.099)	0.051 (0.100)	0.050 (0.099)	0.050 (0.100)	0.050 (0.101)	0.050 (0.099)	0.050 (0.099)
30	0.049 (0.100)	0.049 (0.099)	0.050 (0.099)	0.050 (0.098)	0.051 (0.099)	0.049 (0.101)	0.050 (0.100)	0.050 (0.099)
50	0.048 (0.098)	0.049 (0.099)	0.049 (0.097)	0.049 (0.097)	0.049 (0.100)	0.049 (0.098)	0.048 (0.097)	0.047 (0.097)
75	0.049 (0.098)	0.048 (0.096)	0.048 (0.098)	0.052 (0.102)	0.048 (0.098)	0.050 (0.099)	0.049 (0.099)	0.048 (0.099)
100	0.050 (0.099)	0.049 (0.097)	0.048 (0.098)	0.051 (0.102)	0.049 (0.099)	0.050 (0.099)	0.049 (0.099)	0.049 (0.099)
200	0.049 (0.099)	0.049 (0.098)	0.048 (0.100)	0.049 (0.100)	0.049 (0.097)	0.050 (0.099)	0.048 (0.099)	0.049 (0.098)
300	0.051 (0.097)	0.049 (0.097)	0.049 (0.097)	0.049 (0.098)	0.048 (0.098)	0.049 (0.098)	0.049 (0.099)	0.050 (0.099)
500	0.049 (0.099)	0.050 (0.099)	0.049 (0.097)	0.052 (0.100)	0.049 (0.098)	0.049 (0.097)	0.050 (0.097)	0.050 (0.098)

Table 2. Type I Errors at 0.05(0.10)

Table 2 gives the values for type I errors of the test statistics. There is not much difference in Type I errors. All considered statistics have Type I error values very close to the expected 0.05 and 0.10 significance levels for all considered sample sizes.

3.3. Power Comparison

In this section we present results of power comparisons of univariate normality tests under consideration and analyze results to see which univariate normality tests are the best. We have considered eight univariate normality tests, namely, Z_{ED}, X_{AD}, Gel-Gastwirth, Brys-Hubert-Struyf, Bonett-Seier, Doornik-Hansen, Jaque-Bera, and D'Agostino-Pearson under different alternatives. Four different distribution groups were used with different skewness and kurtosis, namely, symmetric long-tailed, symmetric short-tailed, asymmetric, and normal modified distributions. Each group was tested using Monte Carlo simulations for different sample sizes $10, 15, 20, 30, 50, 75, 100, 200, 300$, and 500 taking $10,000$ samples from each sample size. The significance level used was 0.05.

The Tables 3-11 present simulation results of the considered test statistics of respective normality tests for the four groups for above-mentioned sample sizes with 0.05 significance level. Tables 3 and 4 present results for symmetric long-tailed distributions. Tables 6 and 7 present for symmetric short-tailed, Tables 7-9 present for asymmetric, and Tables 10 and 11 present for normal modified distributions, respectively. Tables 12-15 present the average values for all distributions in the four groups symmetric long-tailed, symmetric short-tailed, asymmetric, and normal modified distributions under consideration based on sample sizes.

Twelve cases of mixture of normal distributions were considered. These mixture distributions, denoted by $MixN(p; a; b)$, consisting of randomly selected observations with probability $1 - p$ drawn from a standard normal distribution and with probability p drawn from a normal distribution with mean a and standard deviation b. Next, we consider five cases of standard normal distributions with outliers and denoted by $Nout(a = 1)$ to $Nout(a = 5)$. This set of distributions was specifically considered in order to identify which normality tests are less sensitive to outlier observations that may be present in an underlying normal data sample. Finally, four cases of the standard normal distribution truncated at a and b $Trunc(a; b)$, where a and b are the lower and upper truncations points, respectively, were considered.

From Tables 3 and 4 for symmetric long-tailed distributions it is evident that the Gel-Gastwirth test (RJB statistic) is the best showing maximal power for almost all sample sizes considered. Table 12 that presents average power values based on sample sizes, as expected, ranks Gel-Gastwirth test as the best test too. The second and third best tests, considering the average of the power for all sample sizes for these symmetric long-tailed distributions, are Doornik-Hansen (E_p) and Jaque-Bera (JB) respectively. However, there is not much difference in power between these three tests. But, if we consider the power of Brys-Hubert-Struyf Test (T_{ML} Statistic), the power is substantially low.

From Tables 5, 6 and 13, for symmetric short-tailed distributions, it is evident that the test Z_{ED} is the best with maximal average power. The second and third best tests are X_{AD} and Bonett-Seier (T_w Statistic) tests.

From Tables 9-11 and 14, for asymmetric distributions, the order of the best three tests with highest average power are Doornik-Hansen (E_p), X_{AD}, and Jaque-Bera (JB), respectively. There is not much difference in power among these tests. In addition, the D'Agostino-Pearson (OT) and the Gel-Gastwirth (RJB) tests do not have much of an average power difference with these three best tests for asymmetric distributions. However, Bonett-Seier (T_w) test has a considerably lower power on average based on considered sample sizes for asymmetric distributions.

For the normal modified distributions, Tables 10-11 and 15 suggest that Doornik-Hansen (E_p), X_{AD}, Jaque-Bera (JB), D'Agostino-Pearson (OT), and Gel-Gastwirth (RJB) are the best, in order (E_p being the best), with not much of a difference in average power among them. The Z_{ED}, Bonett-Seier (T_w), and Brys-Hubert-Struyf (T_{ML}) tests seem to have a comparatively lower average power for normal modified distributions.

Thus, the results suggest that the tests, Gel-Gastwirth (RJB) for symmetric long-tailed, Z_{ED} for symmetric short-tailed, and Doornik-Hansen (E_p) for both asymmetric and normal modified distributions, respectively, are the best in terms of power and work well.

Analyzing all the simulation results for all different sample sizes, the univariate normality test Doornik-Hansen (E_p) can be recommended as the best to test univariate normality when the nonnormality of the distribution is unknown. In addition, values suggest that power increases when sample size increases.

2*n	Test Statistics							
	X_{AD}	Z_{ED}	RJB	T_{ML}	T_w	E_p	JB	OT
10	21.105	18.744	**23.325**	6.140	17.566	22.185	22.065	22.010
15	28.544	26.388	**30.735**	6.994	25.233	29.454	28.790	28.301
20	33.898	31.742	**36.065**	7.522	30.745	34.322	34.151	33.106
30	40.780	39.041	**43.110**	7.995	38.066	41.348	41.176	39.378
50	48.644	47.688	**51.054**	10.785	46.632	49.131	49.077	46.410
75	54.526	54.154	**57.027**	14.898	52.376	55.307	55.285	52.145
100	58.685	58.509	**61.441**	18.302	56.590	59.467	59.738	56.475
200	67.278	67.541	**70.013**	26.130	65.094	68.678	69.123	66.136
300	71.558	72.268	**74.398**	30.998	69.311	73.240	73.776	71.143
500	76.833	78.039	**79.486**	37.295	74.938	78.616	79.005	76.923
ADB	2.48	3.25	0.00	35.95	5.01	1.49	1.44	3.46
Rank	4	5	1	8	7	2	3	6

Table 3. Average Power of Long-tailed Distributions

2*n	Test Statistics							
	X_{AD}	Z_{ED}	RJB	T_{ML}	T_w	E_p	JB	OT
10	14.647	13.344	4.149	**15.470**	12.408	13.068	4.217	6.418
15	19.156	**19.962**	3.716	16.943	18.013	16.501	3.683	12.489
20	23.415	**25.977**	3.693	19.435	23.446	19.325	3.662	19.205
30	30.723	**35.707**	4.228	20.410	32.045	25.335	5.118	31.277
50	42.466	**48.538**	5.311	23.553	43.262	36.393	17.685	46.218
75	52.090	**57.217**	8.154	27.299	52.027	48.310	31.642	49.944
100	57.757	**62.266**	22.891	29.939	57.878	55.062	46.547	51.959
200	69.786	**73.785**	58.998	37.402	69.589	67.860	64.611	58.180
300	76.874	**80.327**	69.496	43.055	76.282	75.605	73.827	61.303
500	84.386	**87.105**	80.642	49.775	83.984	83.878	82.985	67.555
ADB	3.50	0.21	24.50	22.30	3.74	6.50	17.23	10.18
Rank	2	1	8	7	3	4	6	5

Table 4. Average Power of Short-tailed Distributions

2*n	Test Statistics							
	X_{AD}	Z_{ED}	RJB	T_{ML}	T_w	E_p	JB	OT
10	**19.334**	11.816	17.706	12.837	9.166	19.477	19.318	18.062
15	28.749	15.391	24.621	14.965	11.615	**28.889**	26.901	25.249
20	**36.270**	18.427	30.234	20.909	13.909	36.121	32.930	30.984
30	**45.447**	23.442	38.319	24.040	17.784	45.766	41.849	39.673
50	55.595	30.526	49.061	32.875	23.486	**56.326**	52.441	50.386
75	63.683	36.331	57.264	40.489	28.178	**65.002**	60.641	58.924
100	69.381	40.685	63.541	44.649	32.124	**70.751**	66.918	65.359
200	81.051	50.196	79.726	52.745	41.903	**81.721**	80.606	79.659
300	85.998	56.166	85.449	58.160	47.744	**86.118**	85.451	84.904
500	91.432	63.847	90.634	65.797	55.140	**91.811**	90.919	90.994
ADB	0.51	23.53	4.55	21.46	30.10	0.01	2.41	3.79
Rank	2	7	5	6	8	1	3	4

Table 5. Average Power of Asymmetric Distributions

2*n	Test Statistics							
	X_{AD}	Z_{ED}	RJB	T_{ML}	T_w	E_p	JB	OT
10	16.324	12.739	16.534	7.336	11.634	**16.900**	15.844	15.235
15	**23.881**	17.891	22.956	8.806	17.067	23.599	21.631	20.793
20	**31.106**	22.937	28.223	10.771	21.480	30.054	26.199	24.914
30	**40.527**	28.648	34.853	11.971	26.365	39.964	33.742	30.851
50	**51.759**	32.989	42.165	17.811	30.023	51.349	44.219	40.789
75	**59.195**	34.518	48.483	25.366	30.985	59.172	52.793	50.538
100	**63.214**	35.273	54.113	30.549	31.689	63.224	58.537	57.690
200	69.034	36.998	65.182	42.759	33.422	**69.836**	67.417	69.024
300	71.307	38.883	68.869	49.060	35.235	**73.552**	71.560	73.327
500	74.132	42.595	74.094	55.161	39.273	**78.245**	76.983	77.614
ADB	0.77	20.47	5.27	24.86	23.10	0.23	3.93	4.74
Rank	2	6	5	8	7	1	3	4

Table 6. Average Power of Normal Modified Distributions

4. Concluding Remarks

Analysis of the comparison results suggest that X_{AD} has the best accuracy and type I error out of all eight tests considered. The power comparison suggests the most powerful test varies with the different groups of distributions based on the skewness and kurtosis of the distribution and also on the sample sizes. For all sample sizes considered, based on the averages, for the symmetric long-tailed distributions, the order of the best three tests is the Gel-Gastwirth test (RJB), Doornik-Hansen test (E_p), and Jaque-Bera test (JB). For the symmetric short-tailed distributions the order of best three is Z_{ED}, X_{AD}, and Bonett-Seier (T_w). For the asymmetric distributions and for the group with normal modified distributions, the best three, in order, are Doornik-Hansen (E_p), X_{AD}, and Jaque-Bera (JB). Therefore, the general recommendation from the study, according to the nature of nonnormality, the most powerful tests are Gel-Gastwirth test (RJB) for symmetric long-tailed, Z_{ED} for symmetric short-tailed, and Doornik-Hansen (E_p) for asymmetric and normal modified distributions. Based on the comparison results out of all eight tests compared here, to assess the validity of univariate normality, overall, for all four groups of different distributions (when the nonnormality is unknown), Doornik-Hansen (E_p) test seems to be the best fit on average for all sample sizes. However, with the high importance of normality, as more and more new normality tests are being developed, it is of considerable importance to continue to compare these new normality tests with the tests we compared here, in the future as well.

References

[1] Anderson TW., Darling DA. A Test of Goodness of Fit, *Journal of the American Statistical Association*, **49**(268) (1954) 765–769. ISSN 0165-1765. http://dx.doi.org/10.2307/2281537.

[2] Bonett DG., Seier E. A test of normality with high uniform power, *Computational Statistics and Data Analysis.* **40** (2002) 435–445. 10.1016/S0167-9473(02)00074-9.

[3] Brys G., Hubert M., Struyf A. A robust measure of skewness, *Journal of Computational and Graphical Statistics.* **13**(4) (2004) 996–1017. https://www.jstor.org/stable/27594089.

[4] Brys G., Hubert M., Struyf A. Goodness-of-fit tests based on a robust measure of skewness, *Computational Statistics.* **23**(3) (2008) 429–442. 10.1007/s00180-007-0083-7.

[5] Chen L., Shapiro SS. An alternative test for normality based on normalizedspacings, *Journal of Statistical Computation and Simulation.* **53** (1995) 269–288.

[6] D'Agostino R., Pearson ES. Tests for Departure from Normality. Empirical Results for the Distributions of b_2 and $\sqrt{b_1}$. *Biometrika.* **60**(3) (1973) 613–622. www.jstor.org/stable/2335012.

[7] Desgagné A., de Micheaux PL. A powerful and interpretable alternative to the Jarque–Bera test of normality based on 2nd-power skewness and kurtosis, using the Rao's score test on the APD family, *Journal of Applied Statistics.* **45**(13) (2018). 10.1080/02664763.2017.1415311.

[8] Doornik J., Hansen H. An omnibus test or univariate and multivariate normality, *Working Paper*, (1994). https://doi.org/10.1111/j.1468-0084.2008.00537.x.

[9] Gel YR., Gastwirth JL. A robust modification of the Jarque-Bera test of normality, *Economics Letters.* **99**(1) (2008) 30–32. 10.1016/j.econlet.2007.05.022.

[10] Hubert M., Struyf A. A comparison of some new measures of skewness, *Developments in Robust Statistics.* (2003) 98–113.

[11] Jarque CM., Bera AK. A Test for Normality of Observations and Regression Residuals, *International Statistical Review.* **55**(2) (1987) 163–172.

[12] Lilliefors H. On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown, *Journal of the American Statistical Association.* **62**(318) (1967) 399–402. 10.2307/2283970.

[13] Noughabi H., Arghami N.. Testing Normality Using Transformed Data Testing Normality Using Transformed Data, *Communications in Statistics -Theory and Methods.* **42**(17) (2013) 3065–3075. https://doi.org/10.1080/03610926.2011.611604.

[14] Razali N., Wah Y. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests, *Journal of Statistical Modeling and Analytics.* **2**(1) (2011) 21–33.

[15] Romao X., Delgado R., Costa A. An empirical power comparison of univariate goodness-of-fit tests for normality, *Journal of Statistical Computation and Simulation.* **80**(5) (2010) 545–591. 10.1080/00949650902740824.

[16] Royston J. A Simple Method for Evaluating the Shapiro-Francia W' Test of Non-Normality, *Journal of the Royal Statistical Society.* **32**(3) (1983) 297–300. 10.2307/2987935.

[17] Ryan T., Joiner B. Normal Probability Plots and Tests for Normality, *Technical Report.* (1976).

[18] Shapiro S., Wilk M. An Analysis of Variance Test for Normality (Complete Samples), *Biometrika.* **52** (1965) 591–611. 10.2307/2333709.

[19] Shenton LR., Bowman KO. A Bivariate Model for the Distribution of $\sqrt{b1}$ and b2, *Journal of the American Statistical Association.* **72**(357) (1977) 206–211. 10.1080/01621459.1977.10479940.

[20] Thode HC. *Testing For Normality*, Marcel Dekker, Basel, NY. ISBN 0824796136. 10.1201/9780203910894, (2002).

[21] Yap B., Sim C. Comparisons of various types of normality tests, *Journal of Statistical Computation and Simulation.* **12** (1981) 2141–2155. 10.1080/00949655.2010.520163.