On empirical power of univariate normality testsunder symmetric, asymmetric and scaled distributions

Muyombya Solomon Matovu(MSc. Statistics)1
[1]Department of Economics and Statistics, Kampala International University Uganda, Kampala

solomonmmatovu@gmail.com

Abstract: The study aims at conducting an empirical comparison of powers of the univariate normality tests under different distributions to obtain their ranking using a Monte-Carlo simulation for large sample sizes. A total of six normality tests selected. From the Empirical Distribution Function (EDF), Kolmogorov-Smirnov test (Lilliefors correction) and Anderson-Darling normality tests were chosen. From the regression and correlation family of distributions, Shapiro-Wilk and Shapiro-Francia normality tests were chosen. Jaque-Bera and D'Agostino Pearson normality tests were chosen from the moment family. The empirical powers of these normality tests were studied using distributions that are symmetric, asymmetric and scale contaminated normal distributions. Findings show that for symmetric distributions, Kolmogorov-Smirnov normality test is the most powerful test,Anderson-Darling, Shapiro-Wilk, Shapiro-Francia, D'Agostino-Pearson and lastly Jaque-Bera. For asymmetric distributions, Anderson-Darling normality test,Shapiro-Wilk, Kolmogorov-Smirnov, Jaque-Bera, Shapiro-Francia and lastly D'Agostino. For scale contaminated distributions, Kolmogorov-Smirnov is the most powerful test, Anderson-Darling, Shapiro-Francia, Shapiro-Wilk, D'Agostino-Pearson and lastly Jaque-Bera. Thus, regardless of the nature of the distribution given a large sample size, Kolmogorov-Smirnov is the most powerful normality test, Shapiro-Wilk, Shapiro-Francia, Anderson-Darling, Jaque-Bera and lastly D'Agostino-Pearson. The study recommends that for distributions that have short tails like symmetric distributions, correlation/regression-based tests should be used. For long tailed distributions like symmetric distributions,Empirical-based normality tests should be used and moment-based tests should be used if interest is in kurtosis and skewness of the data.

Key Words: Empirical Distribution Function, correlation/regression tests and Moment-based normality test, statistical power.

1 Introduction

Statistical inference, a branch of statistics that draws on the results obtained from the data analysis and descriptive stage of statistics is a very critical step in the data modeling and research process. Consequently, there is a need to create a comprehensive distinction between parametric and non-parametric statistical analysis, since any further inference made on the data is heavily dependent on the assumptions made under these two branches. The parametric approach to data analysis has attracted substantial appeal in the recent years since its results provide the most plausible estimates of population parameters giving rise to an interest in normality studies.

Due to the relative importance and severity of the outcome associated with violation of normality, several measures of normality right from the graphical to the empirical measures have been developed to examine and detect the presence then extent of normality plus any possible remedies. However, the huge number of empirical and numerical measures of normality will often leave any statistics practitioner or student either spoilt for choice or confused about the best tool to apply in testing since there are over 44 available statistical tools.

The legendary statistician, Fisher was the first to discover and derive the first known tests of normality for investigating the distribution of any given data in the instances where the mean and variances of the data under investigation are known. His standard third and fourth moments for testing normality $\sqrt{b_2}$ and $b_2$grew in popularity and usage but were cut short by the fact that they could not be applied in the instances where the means and variances of a given data were unknown.The shortcomings associated with Fisher's tests of Normality were later on modified by other statisticians eventually resulting into newer and better tests of normality widely used in the modern statistical analysis [1].

Following Fisher's invention of the statistical tests of normality, a series of other tests were developed as either modifications of Fisher's original test on normality or as entirely new statistical tests of normality.Some of the documented tests include; D'Agostino-Pearson K^2test of normality developed by Agostino and Pearson in 1973[2].Pearson test of normality [3], Jarque-Bera test of normality [4] Urzua test of normality [5] the Cho and Im test[6] , Bonett-Seie test of normality [7].Other tests of normality include; Bryns-Hubert-Struyf-Bonnet-Seier Joint Test [8], Coins tests of normality [9], [10] and so many other tests of normality.

Due to the relative significance each of these Normality tests accords to the distribution and spread of the data, they have come to be classified as those based on the empirical distribution function like K-S and AD.Those that are classified on the basis of measures of moments like $K^2$and JB.Lastly, those based on regression and correlation like SW and S-F.Each of these statistics has got its own test statistics considered in developing the statistic and were considered in the study [11].The study investigated the most powerful tests of normality from a set of familiar and conventional measures by comparing their empirical powers.

2 Related literature

Shapiro [13] studied nine different tests of normality in a bid to find out which of the nine could give the most powerful results in testing for normality. Shapiro [13] simulated data and studied the empirical power of Kolmogorov-Smirnov Test (K-S), Cramer-von Mises (W2); Shapiro-Wilk Statistic (SW), Anderson-Darling (AD); Chi-square Test (X^2); Durbin (D); $\sqrt{b_2}$, Studentized Range Test (U) and b_1 independently. The results of this study indicate that the Shapiro-Wilk test produced the best results for distributions that were general and symmetric compared to all the other normality tests that he considered in his study. It was also found out that those tests that considered the Empirical Distribution Function showed the least power in his study.

D'Agostino [2] further discovered that if the primary aim was to investigate whether a given distribution is highly symmetric based on the results of its Kurtosis results, then it would be proper to simply use the Kurtosis and Skewness-based tests [14].
Furthermore, Seir [7] also selected several distributions that are asymmetric and moderately skewed, slightly skewed and highly skewed to investigate the behaviour of these tests in extreme conditions. The study revealed that the regression class of normality tests namely; D'Agostino [2], Chen and Shapiro, Shapiro, Zhang and Royston produced the most significant empirical powers.

Univariate statistical analysis explores distinct values in the data set at an individual level other than considering the variables as a set. Univariate analysis will consider a wide range of values considering their measures of central tendency and any form of dispersion that may be observed within that single data set.

Empirical power of a given variable refers an inherent significance testing analysis defined as the probability of a statistical test falling correctly in the region that would have the null hypothesis rejected in the statistical analysis process. An empirical power of 0.5 would imply that if a given experiment is run multiple times, the existing effects will be discovered at least four times in the entire experiment [12].

3 Objective

To find out the most powerful tests of normality among the six popular tests (K-S, AD,K^2, JB, SW and S-F) based on the results of their empirical power.

4 Materials and methods

The study carried out an automated simulation of data following several distributions selected from the symmetric, asymmetric and scale contaminated family of distributionsat 5% level using advanced statistical software.Several sets of distributions that are symmetric for instance$\beta(1,1)$, $N(0,1)$, $t(10)$were considered (Azzalini & Capitanio, 2003).

Several forms of asymmetric distributions for instance, $\Gamma(3,2)$, $\beta(2,5)$ and $chi(10)$were also considered (Forbes *et al*. 2011). Lastly, several forms of contaminated or modified forms of distributions such as $N(3,1)$, $N(6,9)$ and a Normal distribution with outlierswere also considered in the study (Ross, 2004).

Six different tests were considered in this study. They include; Anderson-Darling (AD), Smirnov Kolmogorov (S-K), Shapiro-Wilk (S-W), Jarque - Bera (J-B), Shapiro-Francia (S-F) and D'Agostino-Pearson test(D-A). Basing on the underlying assumptions made in each of these tests, they are grouped into three categories. The first group consists of those classified as following an Empirical Distribution Function (EDF), measures of moments, regression and correlation tests [15].

5 Results

Table 1:Ranking of empirical power under symmetric distributions

Normality test	Empirical power	Ranking
Kolmogorov-Smirnov	90.56347934	1
Anderson-Darling	76.86862293	2
Shapiro-Wilk	69.91870356	3
Shapiro-Francia	68.19249275	4
D'Agostino-Pearson	64.3736	5
Jaque-Bera	61.80359419	6

Source: Solomon (2016).

Result in Table 1shows that the empirical powers associated with each of the selected normality tests under the selected asymmetric, asymmetric and scale contaminated distributions. Using $\beta(1,1)$, N(0,1),$t(10)\Gamma(3,2)$,$\beta(2,5)$, $chi(10)$,N(3,1), N(6,9) and a Normal distribution with outliers, an average ranking of their statistical power was used to perform a case-wise ranking to identify the strongest normality diagnostics.

Table (1) shows that for the symmetric distributions, Kolmogorov-Smirnov normality test is the most powerful test (average power = 90.56). It is closely followed by Anderson-Darling with an average empirical power of 76.86, Shapiro-Wilk with an average power of 69.91. This is then closely followed by Shapiro-Francia (average power = 68.19) then D'Agostino-Pearson (average power = 64.37) and lastly Jaque-Bera (average power = 61.80).

Table 2:Ranking of empirical power under asymmetric distributions

Normality test	Empirical power	Ranking
Anderson-Darling	99.54208	1
Shapiro-Wilk	99.50926	2
Kolmogorov-Smirnov	99.30276	3
Jaque-Bera	99.27596	4
Shapiro Francia	99.01062	5
D'Agostino-Pearson Test	95.25428	6

Source: Solomon [16].

Table 2 shows that for the asymmetric distributions, Anderson-Darling normality test is the most powerful test (average power = 99.54). It is closely followed by Shapiro-Wilk normality test with an average empirical power of 99.51, Kolmogorov-Smirnov with an average power of 99.30. This is then closely followed by Jaque-Bera (average power = 99.28), Shapiro-Francia (average power = 99.01) and lastly D'Agostino (average power = 95.25).

Table 3:Ranking of empirical power under scale contaminated distributions

Normality test	Empirical power	Ranking
Kolmogorov-Smirnov	86.66667	1
Anderson-Darling	63.08533	2
Shapiro-Francia	59.23825	3
Shapiro-Wilk	59.03404	4
D'Agostino-Pearson Test	58.88552	5
Jaque-Bera	58.74429	6

Source: Solomon [16].

Lastly, for the scale contaminated distributions, Kolmogorov-Smirnov normality test is the most powerful test (average power = 86.67). It is closely followed by Anderson-Darling with an average empirical power of 63.08533, Shapiro-Francia with an average power of 59. 23825.This is then closely followed by Shapiro-Wilk (average power = 59.03) then D'Agostino-Pearson (average power = 58.89) and lastly Jaque-Bera (average power = 58.74).

Table 4:Overall empirical powers associated with each distribution in the study.

Distribution	Kolmogorov-Smirnov	Shapiro-Wilk	Shapiro Francia	Anderson-Darling	Jaque-Bera	D'Agostino-Pearson Test
$\beta(1,1)$	98.28	99.72	99.16	99.56	96.41	99.19
$N(0,1)$	86.95	51.68	41.29	58.38	42.29	42.96
$t(10)$	86.45	58.35	64.13	72.66	46.71	50.97
$\Gamma(3,2)$	99.26	99.99	98.52	99.98	99.99	99.98
$\beta(2,5)$	98.85	98.55	98.52	98.68	97.84	85.79
$chi(10)$	99.80	99.99	99.99	99.96	100.00	99.99
$N(3,1)$	80.00	40.34	39.68	55.71	32.37	31.78
$N(6,9)$	80.00	80.00	80.00	74.62	9.48	10.41
OutlierNormal	100.00	100.00	100.00	100.00	100.00	100.00
Average power	**92.18 (1)**	**80.96 (3)**	**80.14 (4)**	**84.40 (2)**	**69.45 (5)**	**69.01(6)**

Source: Solomon [16].

Result in table 4, it can be seen that for the symmetric beta distributions, β(1,1), S-W is the most powerful normality test and A-D is the least powerful though all the tests produced closely powerful results. For the standard normal distribution, N(0,1), K-S produced the best results and S-F produced the least powerful estimates.

Furthermore, for the symmetric T- distribution $t(v = 10)$, K-S produced the results and J-B produced the least but S-W performed poorly possibly due to the long tails. For the asymmetric gamma distribution,$\Gamma(3,2)$,both the S-W and J-B produced the most powerful results. J-B and A-D, produced the best results for the asymmetric $chi(v = 10)$ distribution due to the high level of skewness associated with such a distribution.

Additionally, K-S produced the most powerful results under both $N(3,1)$ and $N(6,9)$ scale contaminated distributions. Lastly, these normality tests produced similar empirical powers for the distributions that are contaminated with outliers with very consistent p-values that also rejected the null hypothesis regardless of the sample size.

Overall, K-S is the most powerful normality test, S-W, S-F, A-D, J-B and lastly D-A as shown in table (4) regardless of the nature of the distributions under consideration based on the empirical power each of these normality tests produced.

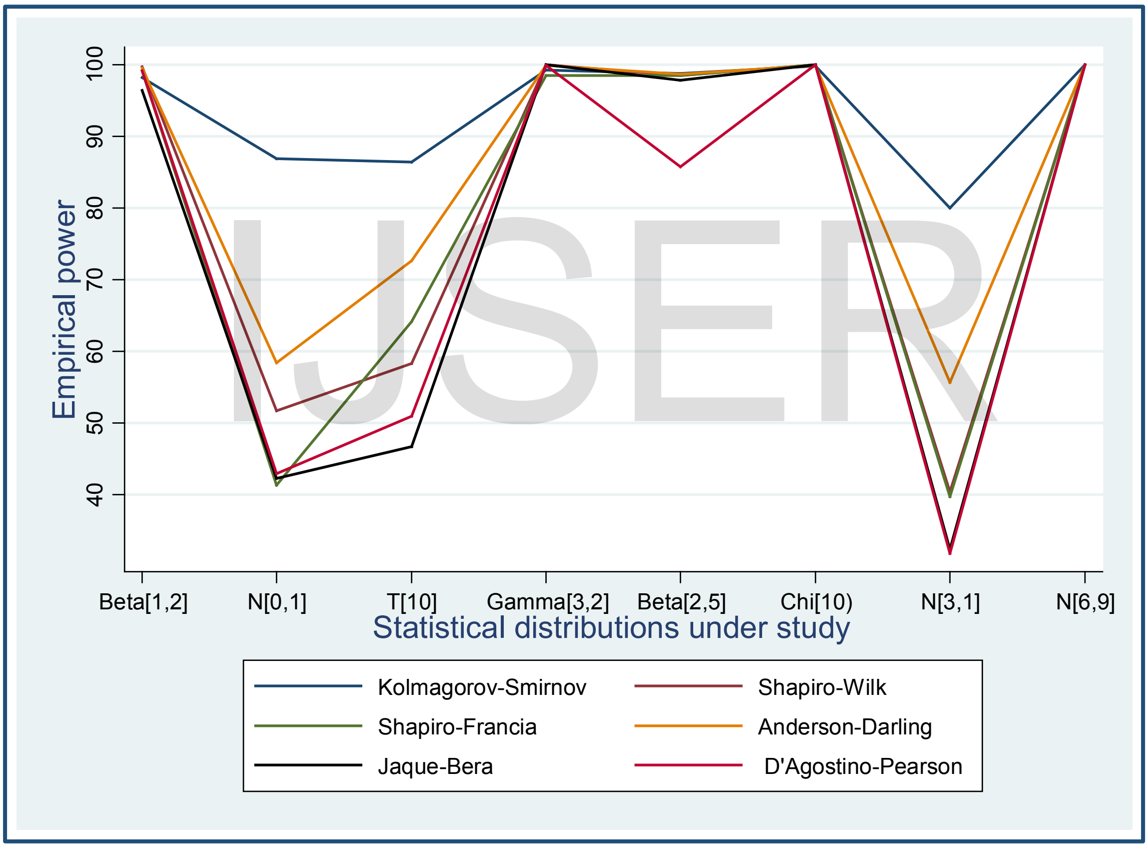

Figure 1: Overall empirical powers associated with each distribution in the study

Source: Solomon [16].

Result in figure 1 presents a summary of the empirical powers associated with each of the normality tests for the different distributions in the study. It presents an overall value for empirical power after consideration of all the distributions.

5 Discussion

The analysis thus revealed that for the symmetric family of distributions, the Empirical Distribution Function based tests have the best statistical power, followed by Regression and Correlation results and lastly those based on the measures of moments. The long tails often associated withsymmetric distributions offers an added advantage to the EDF based normality tests.

For the asymmetric distributions, Empirical Distribution Function based tests have the best statistical power, followed by Regression and Correlation results and lastly those based on the measures of moments for the asymmetric distributions as well.However, the difference between these normality power results were not statistically significant due to the fact that beside the chi-squared distributions, the asymmetric distributions in the study have very short tails on either sides granting both the Regression and Correlation and the moment-based normality tests added strength. Furthermore, due to the low kurtosis and skewness statistics for each of the statistics in the study, the moment-based normality tests are associated with low power since they are heavily reliant on kurtosis and skewness. For the scale contaminated distributions, the moment-based tests produced the best tests due to the huge kurtosis and skewness figures associated with such distributions.

6 Conclusion

The objective of the study was to find out the most powerful tests of normality among the six popular tests (K-S, AD,K^2, JB, SW and S-F) based on the results of their empirical power.After adjusting for symmetry,assymetry and scale contamination, the results indicate that, K-S is the most powerful normality test, S-W, S-F, A-D, J-B and lastly D-A regardless of the nature of the distribution based on the empirical power each of these normality tests produced. That indicates that Empirical-based normality tests are consistently associated with higher empirical power if compared to the correlation/regression and moment-based normality tests.

7 Recommendations

1. For distributions that have relatively long tails like some symmetric distributions, the EDF-based nomality tests ought to be used.
2. For distributions that have relatively short tails like some asymmetric distributions, the correlation/regression based nomality tests ought to be used.
3. If the goal is only to investigate the skewness and kurtosis of the distribution, then the moment-based normality test ought to be used.

References

[1] Henderson, A. R. (2006). Testing experimental data for univariate normality. Clinica Chimica Acta, 366(1), 112-129.

[2] D'Agostino, R. A. L. P. H., & Pearson, E. S. (1973). Tests for departure from normality. Empirical results for the distributions of b2 and√ b1. *Biometrika*, *60*(3), 613-622.

[3] Pearson, E. S., & Bowman, K. O. (1977). Tests for departure from normality: Comparison of powers. *Biometrika*, *64*(2), 231-246.

[4] Jarque, C. M., & Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review/Revue Internationale de Statistique*, 163-172.

[5] Urzúa, C. M. (1996). On the correct use of omnibus tests for normality. Economics Letters, 53(3), 247-251.

[6] Ross, S. M. (2004). *Introduction to probability and statistics for engineers and scientists*. Academic press.

[7] Bonett, D. G., & Seier, E. (2002). A test of normality with high uniform power.*Computational statistics & data analysis*, *40*(3), 435-445.

[8] Brys, G., Hubert, M., & Struyf, A. (2008). Goodness-of-fit tests based on a robust measure of Skewness. Computational statistics, 23(3), 429-442.

[9] Coin, D. (2008). A goodness-of-fit test for normality based on polynomial regression. Computational statistics & data analysis, 52(4), 2185-2198.

[11] Razali, N. M., & Wah, Y. B. (2011). Power comparisons of Shapiro-wilk, Kolmogorov-smirnov, Lilliefors and Anderson-darling tests. Journal of Statistical Modeling and Analytics, 2(1), 21-33.

[12] Brandt, S. (2012). *Data analysis: statistical and computational methods for scientists and engineers*. Springer Science & Business Media.

[13] Shapiro, S. S., Wilk, M. B., & Chen, H. J. (1968). A comparative study of various tests for normality. Journal of the American Statistical Association, 63(324), 1343-1372.

[14] Azzalini, A., & Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *65*(2), 367-389.

[16] Solomon, M. M. (2016). Comparative Empirical Power of Univariate Normality tests: A Case study of the 2016 Uganda General Election Results. Kampala International University, Uganda: Unpublished Research Thesis.

ABOUT THE AUTHOR

Muyombya Solomon Matovu is a Professional Freelance Statistician. He is a member of the Uganda Statistical Society. He holds a Master of Science (M.Sc.) degree in Statistics from Kampala International University, Uganda. His research interests are in Statistical computing, simulation and data visualisation.