Applications of Normality Test in Statistical Analysis

Nasrin Khatun

Department of Statistics, Jahangirnagar University, Savar, Dhaka, Bangladesh
Email: nasrin@juniv.edu

Abstract

In this study, to power comparison test, different univariate normality testing procedures are compared by using new algorithm. Different univariate and multivariate test are also analyzed here. 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. Also to test the randomness of generated random numbers. For this purpose, 1000 data sets with combinations of sample size n = 10, 20, 25, 30, 40, 50, 100, 200, 300 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. Observed that with the increase of n, overall powers are increased but Shapiro Wilk (SW) test, Shapiro Francia (SF) test and Andeson Darling (AD) test are the most powerful test among other tests. Cramer-Von-Mises (CVM) test performs better than Pearson chi-square, Lilliefors test has better power than Jarque Bera (JB) Test. Jarque Bera (JB) Test is less powerful test among other tests.

Keywords

Normality Test, Univariate Test, Multivariate Test

1. Introduction

In parametric analysis assuming that population is normal. In this case, checking whether population is normal or not. Normality tests are used in different sectors. One application of normality tests is to the residuals from a linear regression model. If they are not normally distributed, the residuals should not be used in Z tests or in any other tests derived from the normal distribution, such as t tests, F tests and chi-squared tests. If the residuals are not normally distributed, then the dependent variable or at least one explanatory variable may have the

wrong functional form, or important variables may be missing, etc. Correcting one or more of these systematic errors may produce residuals that are normally distributed.

After testing normality if data are not normal, to apply Box Cox transformation method for transforming non-normality data to normal. So main aim is to discuss whether the goodness of fit of a statistical model describes how well it fits a set of observations. Measures of goodness of fit typically summarize the discrepancy between observed values and expected values under the model in question.

There is the use of size corrected method in the determination of power. Main aim is to propose a new algorithm for testing multivariate normality. A random number generator is a computational or physical device designed to generate a sequence of numbers or symbols. Nowadays, after the advent of computational random number generator, a growing number of government-run lotteries, and lottery games, are using RNGs instead of more traditional drawing methods, such as using ping-pong or rubber balls etc.

If normality is not a durable assumption, then one alternative is to ignore findings of the normality check and proceed as the data are normally distributed. But this is not practically recommended because in many situations it could lead to incorrect calculations. Due to countless possible deviations from normality, Andrews *et al.* [1] concluded that multiple approaches for testing MVN would be needed.

Conover *et al.* [2] observed that the Kolmogorov Smirnov statistic belongs to the supremum class of EDF statistics and this statistics is based on the largest vertical difference between hypothesized and empirical distribution.

Gray *et al.* [3] observed that the multivariate normality of stock returns is a crucial assumption in many tests of assets pricing models. This paper utilizes a multivariate test procedure, based on the generalized method of moments, to test whether residuals from market model regressions are multivariate normal.

Kankainen *et al.* [4] observed that classical multivariate analysis is based on the assumption that the data come from a multivariate normal distribution. Several tests for assessing multinormality, among them Mardia's popular multivariate skewness and kurtosis statistics, are based on standardized third and fourth moments. In that report, they investigate whether, in the test construction, it is advantageous to replace the regular sample mean vector and sample covariance matrix by their affine equivariant robust competitors.

Koizumi *et al.* [5] observed that some tests for the multivariate normality based on the sample measures of multivariate skewness and kurtosis. For univariate case, Jarque and Bera proposed bivariate test using skewness and kurtosis. They propose some new bivariate tests for assessing multivariate normality which are natural extensions of Jarque-Bera test.

Major power studies done by Pearson *et al.* [6] have not arrived at a definite answer but a general consensus has been reached about which tests are powerful.

Richardson *et al.* [7] observed that a general procedure that takes account of correlation across assets that focus on both the marginal and joint distribution of

the returns. They find highly significant evidence that residuals are non-normal.

Major power studies done by Shapiro *et al.* [8] have not arrived at a definite answer but a general consensus has been reached about which tests are powerful.

The performance of different univariate normality testing procedures for power comparison are compared by using the new algorithm and different univariate and multivariate test are 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. Also to test the randomness of generated random numbers. Different datasets are generated from uniform distribution and tested by using different tests for randomness. And data were also generated from multivariate normal distribution to compare the performance of power of univariate test by using different new algorithms.

2. Materials and Methods

To complete this study 1000 data sets with combinations of sample size n = 10, 20, 25, 30, 40, 50, 100, 200, 300 were generated from uniform distribution and tested by using different tests for randomness. And data were also generated from multivariate normal distribution to compare the performance of power of univariate test by using different new algorithms. The test statistics are calculated for univariate and multivariate data sets and compare to the appropriate critical values in order to estimate the proportion of rejections of randomness test in each situation, consider that $\alpha = 0.05$. The goodness of fit of a statistical model describes how well it fits a set of observations e.g. to test for normality of residuals, to test whether two samples are drawn from identical distributions, or whether outcome frequencies follow a specified distribution. The purposes of goodness of fit test are to compare an observed distribution to an expected distribution. In assessing whether a given distribution is suited to a data set, the Anderson darling test, Shapiro-Wilk test, Pearson chi-square test, Kolmogorov Smirnov test, Akaike information criterion, Hosmer-Lemeshow test, Cramer-von misses criterion, likelihood ratio test etc. are used and their underlying measures of fit can also be used. The normality test is one of the most important tests among the GOF tests. For testing normality whether a given distribution is suited to a data set, to compare a histogram of the sample data to a normal probability curve, graphical tool, Q-Q plot are used. Simulating univariate random number and test whether normality or not. The Q-Q plot for the foregoing data, which is a plot of the ordered data x_j against the normal quantiles q_j, is shown in the following Figure 1.

From the figure we see that the pairs of points (q_j, x_j) very nearly along a straight line and we would not reject the notion that these data are normally distributed with sample size n = 40.

2.1. Univariate Normality Test Procedure

The purpose of this study is to focus on general goodness of fit tests and their
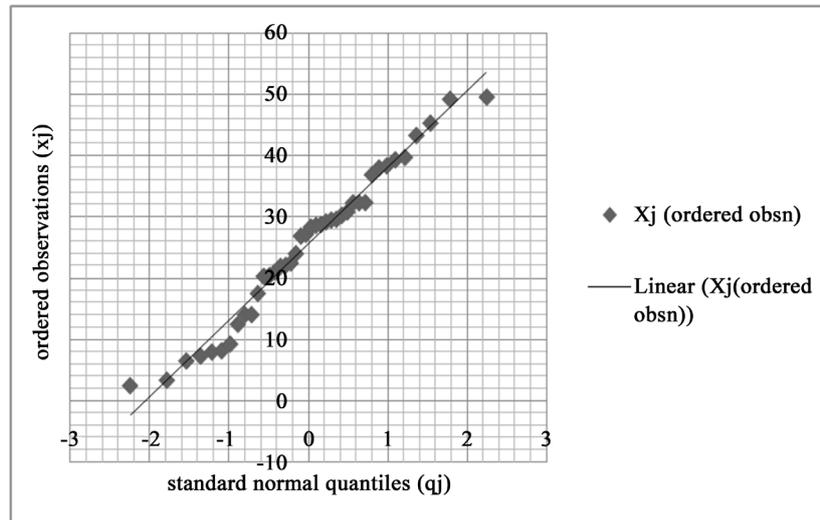
Figure 1. Q-Q plot for the univariate data.

applications to test for normality. While we attempt to mention as broad range of tests as possible, we will be concerned mainly with those tests that have been shown to have decent power at detecting normality to decide which test is appropriate at which situation. In Chi-square test, a single random sample of size n is drawn from a population with unknown cdf F_x. The test criterion suggested by Pearson *et al.* [6] is the random variable $\chi^2 = \sum_{i=1}^{k} \left(\frac{(f_i - e_i)}{e_i} \right)$. A large value of χ^2 would reflect an incompatibility between observed and expected frequencies, and therefore the null hypothesis on which the e_i were calculated should be rejected for χ^2 large. The Shapiro-Wilk test is a test of normality in frequents statistics. The null-hypothesis of this test is that the population is normally distributed. Thus if the *p*-value is less than the chosen alpha level, then the null hypothesis is rejected and there is evidence that the data tested are not from a normally distributed. In other words, the data are not normal. On the contrary, if the *p*-value is greater than the chosen alpha level, then the null hypothesis that the data came from a normally distributed and population cannot be rejected. The Kolmogorov Smirnov statistic belongs to the supremum class of EDF statistics and this statistics is based on the largest vertical difference between hypothesized and empirical distribution Conover *et al.* [2]. The corresponding test reject Ho if $KS \geq K_{1-} \propto$ Cramer-von Misses Test (CM) test is used for judging the goodness of fit of a cumulative distribution function F^* compared to a given empirical distribution function. It is also used as a part of other algorithms, such as minimum distance estimation. If this value is greater than the tabulated value then the hypothesis that the data come from the distribution F can be rejected. Anderson-Darling Test is a modification of the Cramer von Misess Test by introducing appropriate non-negative weight functions $W\left(F_o\left(x\right)\right)$ in order to give different weights to the difference $\left|F_n\left(x\right) - F_o\left(x\right)\right|$. It is a statistical test of whether a given sample of data is drawn

from a given probability distribution. The AD test statistic belongs to the quadratic class of EDF test statistic in which it is based on the squared difference. The corresponding test rejects Ho if $A_n^2 \geq A_{n,\alpha}^2$. The Hosmer Lemeshow test is a statistical test for GOF for logistic regression models. The Kuiper test is used to test whether a given distribution, or family of distributions is contradicted by evidence from a sample of data. The trick with Kuiper test is to use the quantity $D^+ + D^-$ as the test statistic. This small change makes Kuiper test as sensitive in the tails as at the median and also makes it invariant under cyclic transformations of the independent variable. The corresponding test rejects Ho if $V \geq k_{1-\alpha}^{(n)}$. The Akaike information criterion is a measure of the relative quality of statistical models for a given set of data. It provides for model selection. Suppose that we have statistical model for a set of data. Let L be the maximized value of the likelihood function for the model, k be the number of estimated parameters in that model, then the Akaike value of the model is defind by $AIC = 2k - 2\ln(L)$. Given a set of candidate models for the data, the preferred model is one with the minimum AIC value.

2.2. Multivariate Test Procedure

The accessible tests of multivariate normality are procedures based on graphical plots and correlation coefficients, goodness-of-fit tests, tests based on measures of skewness and kurtosis, consistent and invariant test. Our main objective is to develop an implementable algorithm to calculate size corrected powers of competitive MVN tests that can be used to compare the efficiency of various multivariate normality tests. Numerous procedures have been proposed for assessing multivariate normality, some of are the generalized method of moments (GMM), estimating the joint distribution of univariate test statistics, utilizing cross-moments in test procedures, the multivariate jarque-bera test, multivariate extensions of skewness and kurtosis, the multivariate omnibus test, extension of jarque-bera test, omnibus k^2 statistic test, transformed skewness and kurtosis test, modified multivariate jarque-bera test, mardia's mvn test, henze-zirkler's multivariate normality test, royston's multivariate normality test etc.

3. Results and Discussions

3.1. Analysis of Univariate Normality Test

Null hypothesis: Data follows normal distribution.

Alternative hypothesis: Data does not follow normal distribution (Table 1).

From Table 2 observe that p-values for these tests are less than level of significance 0.05. To accept the null hypothesis *i.e.* the data may not come from normal distribution.

3.2. Comparative Study for Generated Random Number of Different Univariate Normality Test

In many power comparison problems, it was not clearly under stable from their

Table 1. Univariate data (radiation data) Richardson *et al*. [7].

Oven no.	radiation	Oven no.	radiation
1	0.15	22	0.05
2	0.09	23	0.03
3	0.18	24	0.05
4	0.1	25	0.15
5	0.05	26	0.1
6	0.12	27	0.15
7	0.08	28	0.09
8	0.05	29	0.08
9	0.08	30	0.18
10	0.1	31	0.1
11	0.07	32	0.2
12	0.02	33	0.11
13	0.01	34	0.3
14	0.1	35	0.02
15	0.1	36	0.2
16	0.2	37	0.2
17	0.02	38	0.3
18	0.1	39	0.3
19	0.01	40	0.4
20	0.4	41	0.3
21	0.1	42	0.05

Table 2. Calculated test statistic by several univariate test.

Test	Test statistic	p-value
jb	22.73	1.16E05
SW	0.87214	0.00023
pearson chi-square	6.8571	0.07659
sf	0.87607	0.00061
lillie	0.21499	3.96E05
cvm	0.31103	0.00023
ad	40.152	1.43E05

power calculation that was their null and alternative hypothesis. That's why we have suggested a new approach for calculating power in our problem, which is size corrected power because of its convenience for easily understanding the power of the null hypothesis and alternative hypothesis as well as for power

comparison with other powers. According to hypothesis, calculate the power and then compare it. It is obvious that the power of all tests under the null hypothesis will be 0.05 but under the alternative hypothesis, the power will be greater because of the greater deviation from null or departure from normality and a particular time it will be closer to 0 or 1. The higher the distance from null hypothesis, higher the power. The hypotheses exhibit the variation of power because of the sample size such as low cell frequency due to small sample size, contamination and change of parameter value. So it's better and convenient to use size corrected power (Table 3).

Figure 2 is concerned with power curves. From Figure 2 and Table 3 see that Shapiro Wilk (SW), Shapiro Francia (SF), Andeson Darling (AD) test are the most powerful test among other tests. Cramer Von Mises (CVM) test perform better than Pearson chi square, Lilliefors and Jarque Bera (JB), Pearson chi-square has better power than Lilliefors and Jarque Bera (JB). Lilliefors test has better power than Jarque Bera (JB). Jarque Bera (JB) is less powerful test among other tests. This study found that no single test has the most powerful in every situation.
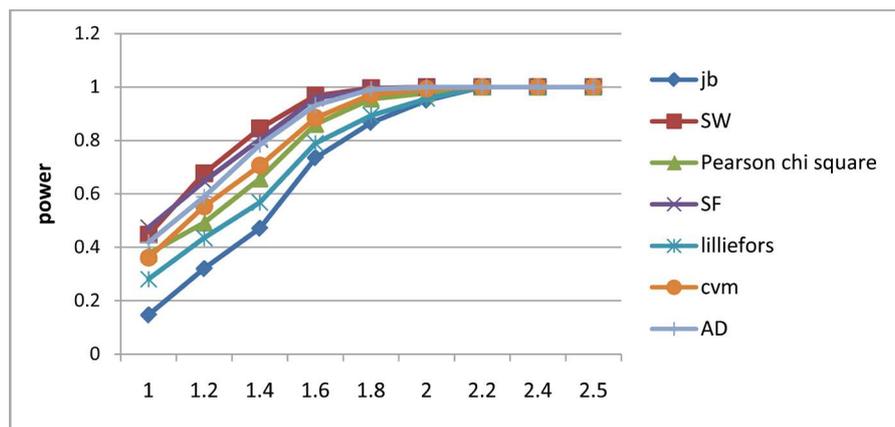

Figure 2. Empirical powers of the univariate normality tests.

Table 3. Powers of the different values for different sample size.

jb	SW	pearson chi-square	sf	lillie	cvm	power AD
0.147	0.448	0.376	0.474	0.281	0.361	0.419
0.32	0.677	0.492	0.647	0.435	0.553	0.589
0.471	0.845	0.657	0.804	0.568	0.706	0.784
0.735	0.968	0.861	0.958	0.789	0.884	0.932
0.867	0.996	0.956	0.994	0.894	0.972	0.992
0.95	1	0.977	0.999	0.957	0.993	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1

Where, JB = Jarque Bera, AD = Andeson Darling, CVM = Cramer-Von-Mises, SW = Shapiro Wilk (SW), SF = Shapiro Francia (SF), Lilli = Lilliefors.

3.3. Analysis of Multivariate Normality Test by Several Test Statistics

From Table 4 and Table 5 observe that p-values for this test are less than level of significance 0.05. So we may not accept the null hypothesis that is the data may not come from multivariate normal distribution.

Table 4. Four measurements on stiffness, Chapter Name—The multivariate normal distribution, Richardson *et al.* [7].

obsn no.	x1	x2	x3	x4	obsn no.	x1	x2	x3	x4
1	1889	1651	1561	1778	16	1954	2149	1180	1281
2	2403	2048	2087	2197	17	1325	1170	1002	1176
3	2119	1700	1815	2222	18	1419	1371	1252	1308
4	1645	1627	1110	1533	19	1828	1634	1602	1755
5	1976	1916	1614	1883	20	1725	1594	1313	1646
6	1712	1712	1439	1546	21	2276	2189	1547	2111
7	1943	1685	1271	1671	22	1899	1614	1422	1477
8	2104	1820	1717	1874	23	1633	1513	1290	1516
9	2983	2794	2412	2581	24	2061	1867	1646	2037
10	1745	1600	1384	1508	25	1856	1493	1356	1533
11	1710	1591	1518	1667	26	1727	1412	1238	1469
12	2046	1907	1627	1898	27	2168	1896	1701	1834
13	1840	1841	1595	1741	28	1655	1675	1414	1597
14	1867	1685	1493	1678	29	2326	2301	2065	2234
15	1859	1649	1389	1714	30	1490	1382	1214	1284

Table 5. Multivariate normality test by several test statistics.

Test	Test statistic	p value
Energy test	1.4717	0.001
kurtosis	35.381	5.98E05
skewness	19.702	0.00057
Mardia's	7.536804	0.00197
Henze Zirkler	1.038131	0.00596
Royston	9.823858	0.00953

4. Conclusion

Normality test is an important aspect in econometrics or statistical analysis because statistical or econometric model are based on normality test. It is very essential to test the randomness of a random number. A generalization of the familiar bell shaped normal density to several dimensions plays a fundamental role in multivariate analysis. It is based on the assumption that data were generated

from multivariate normal distribution. It is mathematically tractable and gives nice results. The normality of the data, which is a key assumption for making valid inferences, can be tested using various statistical tests or visual inspection. There is no single statistical tool to assess the normality that is as powerful as a well-chosen graph. Assessing the normality using graphical methods does lack objectivity which is not the case when dealing with statistical tests. However, the assessment of normality using statistical tests is sensitive to the sample size. In case of small samples, the null hypothesis of normality is often not rejected; hypothesis of normality is rejected even for small violations. So, the graphical methods should be used to analyze the violation of normality in the light of sample size. In sum, combining graphical methods and test statistics will definitely improve our judgment on the normality of the data. While univariate tests of normality are commonly employed, they are unreliable since they fail to accommodate cross-correlations between test statistics. Observed that with the increase of n, overall powers increasing but Shapiro Wilk (SW), Shapiro Francia (SF), Andeson Darling (AD) test are most powerful test among other tests. Cramer-Von-Mises (CVM) test perform better than Pearson chi-square, Lilliefors and Jarque Bera (JB), Pearson chi-square has better power than Lilliefors and Jarque Bera (JB). Lilliefors test has better power than Jarque Bera (JB) Test. Jarque Bera (JB) Test are less powerful test among other tests. This study found that no single test has the most powerful in every situation. In multivariate test analysis we may not accept the null hypothesis that is the data may not come from multivariate normal distribution. There are three popular test statistics for testing multivariate normality test whose are Henze-Zirklers's test, Mardia's Test and Royston's multivariate normality test respectively. From the graphical method chi-square Q-Q plot we can observe that the data are not scattered around the 45-degree line with a positive slope, the greater the departure from this line, the greater the evidence for the conclusion that the series is not normally distributed. Here the data departures from the line so we may conclude that the series is not normally distributed.

5. Limitations of the Study

In this case, we use R programming language which is not always best. Another limitation on sample size and dimension is imposed to keep the amount of computing time reasonable and to consider sample size that minimal for the multivariate analysis. These small sample sizes are likely to be the situation where the assumption of the multivariate normality is most critical to the researchers.

Acknowledgements

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Andrews, D., Gnanadesikan, R. and Warner, J. (1973) Methods for Assessing Multivariate Normality. In: Krishnaiah, P.R., Ed., *Proceedings of the International Symposium on Multivariate Analysis*, **3**, 95-116. https://doi.org/10.1016/B978-0-12-426653-7.50012-0

[2] Conover, W.J. (1999) Practical Nonparametric Statistics. Third Edition, John Wiley & Sons, New York.

[3] Gray, P., Kalotay, E. and McIvor, J. (1998) Testing the Multivariate Normality of Australian Stock Returns. *Australian Journal of Management*, **23**, 135-150. https://doi.org/10.1177/031289629802300201

[4] Kankainen, A., Taskinen, S. and Oja, H. (2003) On Mardia's Tests of Multinormality. *Conference Paper*, 2 May 2003.

[5] Koizumi, K., Okamoto, N. and Seo, T. (2009) On Jarque-Bera Tests for Assessing Multivariate Normality. *Journal of Statistics: Advances in Theory and Applications*, **1**, 207-220.

[6] Pearson, E.S., D'Agostino, R.B. and Bowman, K.O. (1977) Tests for Departure from Normality: Comparison of Powers. *Biometrika*, **64**, 231-246. https://doi.org/10.1093/biomet/64.2.231

[7] Johnson, R.A. and Wichern, D.W. (2015) Applied Multivariate Statistical Analysis, 6th Edition, Pearson, India.

[8] Shapiro, S.S., Wilk, M.B. and Chen, H.J. (1968) A Comparative Study of Various Tests of Normality. *Journal of the American Statistical Association*, **63**, 1343-1372. https://doi.org/10.1080/01621459.1968.10480932