International Journal of Probability and Statistics 2018, 7(5): 130-137
DOI: 10.5923/j.ijps.20180705.02

Comparison of Some Common Tests for Normality

Ogunleye L. I.[1,*], Oyejola B. A.[2], Obisesan K. O.[3]

[1]Business Process Reengineering, Guaranty Trust Bank Plc. Plot 1400, Tiamiyu Savage Street, Lagos, Nigeria
[2]Department of Statistics, University of Ilorin, Ilorin, Nigeria
[3]Department of Statistics, University of Ibadan, Ibadan, Nigeria

Abstract The normal distribution is the bedrock of many statistical procedures. Inferences and conclusions from parametric statistical analysis may not be valid when the normality assumption is violated. Three common procedures used for evaluating whether a random sample of independent observations come from a population with normal distribution are: graphical methods (histograms, box plots, Q-Q-plots), numerical methods (skewness and kurtosis) and formal normality tests. In this study, the type I error rates and power of four common formal tests of normality: Anderson-Darling (AD) test, Chi-square (CS) test, Kolmogorov-Smirnov (KS) test and Shapiro-Wilk (SW) test were compared. Type I error rate of the four tests were computed via simulation (in R) of sample data generated from the standard normal while power comparisons was conducted using common continuous and discrete type as well as less common mixture normal alternative distributions. Five thousand independent samples of various sample sizes were generated from the different distributions considered. Our findings reveal that Shapiro-Wilk test has the most acceptable type I error rate amongst the four tests, followed by Kolmogorov-Smirnov test, Anderson-Darling test and Chi-square test. The power study revealed that none of the four tests is uniformly most powerful for all types of alternative distributions under consideration. Shapiro-Wilk test is the most powerful amongst the four normality tests for continuous –type alternative distributions while Chi-square test outperforms the other three tests for discrete-type distributions. All four normality tests have significantly low powers under the mixture normal distributions with unequal means and equal variances irrespective of the mixture probabilities while there is an improved performance under mixture normals with unequal means and unequal variances.

Keywords Simulation, Normality tests, Mixture Normal, Type I error rates, Power of test

1. Introduction

According to Thode (2002), "normality is one of the most common assumptions made in the development and use of statistical procedures." The dependence of most parametric statistical methods on the normality assumption shows the importance of normality tests in statistical analysis. Inferences from parametric statistical analysis may not be valid when the normality assumption is violated. Therefore, before embarking on any statistical analysis, it is important to test the normality assumption. The easiest way to assess normality is by using graphical methods. The normal quantile-quantile plot (Q-Q plot) is one of the most effective and commonly used diagnostic tool for checking normality of data. Even though the graphical methods are useful tool in assessing normality, they do not provide not sufficient conclusive evidence that the normality assumption holds. Therefore, to support the graphical methods, formal procedures namely numerical methods and normality tests should be performed before passing any judgement about the normality of the data.

The numerical methods include skewness and kurtosis coefficients whereas normality test is a more formal hypothesis testing procedure to ascertain if a particular data follows a normal distribution or not. Significant number of normality tests are available in literature, however, the most common normality test procedures available in statistical software packages are the Anderson-Darling (AD) test, Chi-square (CS) test, Jarque-Bera (JB) test, Kolmogorov-Smirnov (KS) test, Lilliefors test and Shapiro-Wilk (SW) test. The different tests of normality often generate different outputs i.e. some tests reject while others accept the same the null hypothesis of normality for the same dataset. These conflicting results could be misleading and often confusing. Therefore, the choice of test of normality to be used under different circumstance should be given significant attention. Consequently, this study seeks to explain the behaviours of four normality tests under selected discrete and continuous distributions.

* Corresponding author:
lateefogunleye1@gmail.com (Ogunleye L. I.)
Published online at http://journal.sapub.org/ijps

2. Methodology

TYPE *I* ERROR RATE and POWER

The most important property of a test is that it guarantees that the rate of erroneously rejecting the null hypothesis will not be too high. Otherwise, we would reject H_0 too often when in fact the sample comes from a normal population.

The test decision of rejecting the null hypothesis when it is actually true is called type I error. The probability of making a type I error is denoted α and often called the significance level of a test.

We performed empirical studies based on simulations. The underlying principle is described as follows:

If N is the number of randomly generated independent samples of size n where all N samples follow a standard normal distribution, the empirical type I error rate $\alpha_{n,N}$ of a given test for a given sample size n is given by

$$\alpha_{n,N} = \frac{r}{N}$$

Where r is the number of times the null hypothesis is rejected in N tests. The objective is to check for a given test of normality if α_n is higher or lower than α.

The test decision of not rejecting the null hypothesis H_0 when the alternative hypothesis H_1 is true is called type II error and is denoted by β. The power of a statistical test (1- β) is the probability of making the right decision i.e. rejecting a null hypothesis when it is not true.

Let N be the number of randomly generated independent samples of sizes n, where all N samples follow the same distribution that is non-normal. The empirical power $1-\beta_{n,\alpha,N}$ of a given test for normality for a given significance level α is given by

$$1-\beta_{n,\alpha,N} = \frac{m}{N},$$

Where m ≤ N is the number of the m tests that reject the null hypothesis of a normally distributed sample at the significance level α.

In this study, simulation procedure was used to evaluate the empirical power of AD, CS, KS and SW test statistics in testing if a random sample of n independent observations come from a population with normal $N(\mu, \sigma^2)$ distribution. The null and alternative hypotheses are:

$$H_0: The\ distribution\ is\ normal$$

$$H_1: The\ distribution\ is\ not\ normal$$

Three levels of significance α = 1%, 5% and 10% were considered to investigate the effect of the significance level on the power of the tests. In order to obtain the simulated power of the four of the four normality tests, the setting for the simulation parameters were the same as in the empirical type II error investigations.

VALUES OF PARAMETERS USED IN THE STUDY

n = 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 1000
N = 5000
α = 0.01, 0.05 and 0.10

Normality Tests: Anderson-Darling (AD), Chi-square (CS), Kolmogorov-Smirnov (KS), and Shapiro-Wilk (SW)

The alternative distributions were selected to cover both continuous and discrete probability distributions. The alternative distributions considered were three continuous distributions; U (0, 1), Beta (2, 2), Gamma (4, 5) and two discrete distributions; Binomial (n, 0.5) and Poisson (4).

3. Results & Discussion

Table 1 shows the result of the rate at which each of the four normality tests reject a true null hypothesis.

3.1. Type I Error

Table 1. Empirical type I error rate for each test and given sample size

	Sample size n					
Test	10	20	30	40	50	100
AD	0.0498	0.0464	0.0502	0.0526	0.0540	0.0414
CS	0.0648	0.0522	0.0544	0.0590	0.0526	0.0466
KS	0.0468	0.0458	0.0532	0.0530	0.0524	0.0488
SW	0.0504	0.0494	0.0512	0.0504	0.0576	0.0438
Test	200	300	400	500	1000	
AD	0.0508	0.0474	0.0452	0.0468	0.0498	
CS	0.0536	0.0622	0.0488	0.0492	0.0498	
KS	0.0530	0.0496	0.0460	0.0492	0.0498	
SW	0.0496	0.0480	0.0456	0.0508	0.0504	

In order to compare the performance of the different normality tests, the ranking procedure was applied. Rank 1 was assigned to the test whose type I error rate is closest to 0.05 while rank 4 was given to the test that is least close to 0.05. The ranking was done for each sample size and the sum of ranks for each test was obtained. The test with the lowest sum of rank is considered as the best test among those in our collection. Table 2 shows the sum of ranks across all sample sizes for each test.

Table 2. Sum of ranks based on type I error rates for each normality test

Test	Sum of ranks
AD	29
CS	31.5
KS	25.5
SW	24

Table 2 shows that the Shapiro-Wilk test is the best among the four normality tests because it has the lowest sum of ranks. It is closely followed by the Kolmogorov-Smirnov test and the Anderson-Darling test. Chi-square goodness-of-fit test has the largest sum of ranks and hence the poorest.

The type I error rate of each of the four tests does not show any consistent pattern with changes in the sample size.

3.2. Power of the Tests

3.2.1. Power for Continuous Alternative non-normal Distributions

The following tables show that the power of the tests varies with the significance levels, sample sizes and three different continuous alternative distributions. The tables show the power for selected alternative distributions for α = 0.01, 0.05 and 0.10.

Table 3. Power Comparison for Different Normality Tests against U (0, 1) alternative distribution at α = 0.01

	Simulated Power of test			
	α = 0.01			
n	**AD**	**CS**	**KS**	**SW**
10	0.0134	0.0266	0.0232	0.0128
20	0.0364	0.0254	0.0244	0.0252
30	0.0900	0.0354	0.0334	0.0906
40	0.1694	0.0564	0.0398	0.2054
50	0.2666	0.0788	0.0460	0.3588
100	0.7946	0.2760	0.0822	0.9464
200	0.9988	0.7920	0.1938	1.0000
300	1.0000	0.9862	0.3602	1.0000
400	1.0000	0.9996	0.5214	1.0000
500	1.0000	1.0000	0.6826	1.0000
1000	1.0000	1.0000	0.9898	1.0000

Table 4. Power Comparison for Different Normality Tests against U (0, 1) alternative distribution at α = 0.05

	Simulated Power of test			
	α = 0.05			
n	**AD**	**CS**	**KS**	**SW**
10	0.0866	0.0902	0.0834	0.0858
20	0.1634	0.0798	0.0996	0.1884
30	0.2974	0.1058	0.1204	0.3818
40	0.4384	0.1658	0.1384	0.5742
50	0.5782	0.1898	0.1528	0.7476
100	0.9530	0.4550	0.2486	0.9970
200	1.0000	0.9000	0.4740	1.0000
300	1.0000	0.9958	0.6854	1.0000
400	1.0000	0.9980	0.8386	1.0000
500	1.0000	1.0000	0.9290	1.0000
1000	1.0000	1.0000	0.9996	1.0000

Table 5. Power Comparison for Different Normality Tests against U (0, 1) alternative distribution at α = 0.10

	Simulated Power of test			
	α = 0.10			
n	**AD**	**CS**	**KS**	**SW**
10	0.1620	0.1454	0.1498	0.1750
20	0.2844	0.1828	0.1738	0.3542
30	0.4448	0.2178	0.2094	0.5706
40	0.5962	0.2592	0.2322	0.7522
50	0.7232	0.2900	0.2566	0.8758
100	0.9828	0.5770	0.3908	0.9996
200	1.0000	0.9406	0.6462	1.0000
300	1.0000	0.9982	0.8310	1.0000
400	1.0000	0.9998	0.9304	1.0000
500	1.0000	1.0000	0.9816	1.0000
1000	1.0000	1.0000	1.0000	1.0000

Table 6. Power Comparison for Different Normality Tests against Beta (2, 2) alternative distribution at α = 0.01

	Simulated Power of test			
	α = 0.01			
n	**AD**	**CS**	**KS**	**SW**
10	0.0070	0.0118	0.0138	0.0060
20	0.0098	0.0136	0.0160	0.0052
30	0.0122	0.0144	0.0222	0.0072
40	0.0220	0.0138	0.0222	0.0150
50	0.0270	0.0168	0.0182	0.0202
100	0.1086	0.0384	0.0178	0.1416
200	0.3950	0.0942	0.0430	0.6610
300	0.7100	0.1692	0.0630	0.9538
400	0.8998	0.3142	0.0950	0.9978
500	0.9736	0.4692	0.1242	1.0000
1000	1.0000	0.9560	0.3388	1.0000

Table 7. Power Comparison for Different Normality Tests against Beta (2, 2) alternative distribution at α = 0.05

	Simulated Power of test			
	α = 0.05			
n	**AD**	**CS**	**KS**	**SW**
10	0.0452	0.0606	0.0668	0.0440
20	0.0572	0.0512	0.0768	0.0530
30	0.0780	0.0610	0.0742	0.0798
40	0.1088	0.0808	0.0848	0.1150
50	0.1352	0.0738	0.0850	0.1580
100	0.3214	0.1112	0.1068	0.4630
200	0.7110	0.2234	0.1564	0.9250
300	0.9154	0.3786	0.2190	0.9978
400	0.9862	0.5420	0.2816	0.9996
500	0.9976	0.6978	0.3682	1.0000
1000	1.0000	0.9890	0.7050	1.0000

Table 8. Power Comparison for Different Normality Tests against Beta (2, 2) alternative distribution at α = 0.10

	Simulated Power of test			
	α = 0.10			
n	**AD**	**CS**	**KS**	**SW**
10	0.0934	0.1122	0.1190	0.0948
20	0.1234	0.1330	0.1340	0.1246
30	0.1584	0.1314	0.1408	0.1698
40	0.1910	0.1320	0.1412	0.2260
50	0.2348	0.1298	0.1440	0.2906
100	0.4768	0.1944	0.1776	0.6420
200	0.8404	0.3510	0.2740	0.9726
300	0.9668	0.4980	0.3562	0.9998
400	0.9946	0.6644	0.4542	1.0000
500	1.0000	0.7988	0.5250	1.0000
1000	1.0000	0.9970	0.8398	1.0000

For all significance levels considered, the four tests showed very low power values at samples sizes between 10 and 50 inclusive. None of these tests produced power value of 40% at n = 10, 20, 30, 40 and 50. The SW test performed as the most powerful at these sample sizes. At $\alpha = 0.10$ and a pre-determined power of 80%, the SW test requires n > 200 to attain the desired power while a sample size of n > 300 is needed by the AD test be 80% powerful. The CS test will require a sample size that is significantly greater than 500 before it can attain 80% power while a sample size of 1,000 is not large enough for the KS test to attain 80% power.

Tables 6, 7, and 8 above showed that the power behaviours of the four normality tests is different from what we observed under the U (0, 1) alternative distribution. For all significance levels considered, the four tests showed very low power values at samples sizes between 10 and 50 inclusive. None of the tests produced power value of 40% at n = 10 to 50 and the SW test is still the most powerful at these sample sizes.

At α = 0.01 and a pre-determined power of 80%, the SW test requires n > 200 to attain the desired power while a sample size of n > 300 is required by the AD test be 80% powerful. The CS test will require a sample size that is significantly greater than 500 before it can attain 80% power while a sample size as large as 1000 is not large enough for the KS test to be 80% powerful.

For α = 0.05 and a pre-specified power of 80%, the sample sizes required by individual test reduced compared to α = 0.01. The SW test will require n > 100 while a sample size between 200 and 300 will make the AD test 80% powerful. The CS test requires sample size more than 500 to attain 80% power while the KS could not attain this power value even at n= 1000. When α is increased to 0.10, the KS test attain a power value of 83.98% at $n = 1{,}000$ whereas the other tests required lower sample sizes for the desired power.

The four normality tests showed improved power behaviours under the gamma alternative distribution in tables 6, 7 and 8 above. The least powerful KS test in our collection even showed acceptable power values though not comparable to the other three. The power of the SW test at sample size n = 50 for all significance levels considered is greater than 40%. The AD test also showed improved power for these small sample sizes particularly at α = 0.05 and 0.10 respectively.

Table 9. Power Comparison for Different Normality Tests against Gamma (4, 5) alternative distribution at α = 0.01

	Simulated Power of test			
	α = 0.01			
n	**AD**	**CS**	**KS**	**SW**
10	0.0384	0.0258	0.0001	0.0438
20	0.1066	0.0474	0.0001	0.1262
30	0.1848	0.0724	0.0002	0.2358
40	0.2720	0.0790	0.0006	0.3548
50	0.3648	0.1114	0.0014	0.4736
100	0.7372	0.2840	0.0054	0.8616
200	0.9858	0.6842	0.0710	0.9984
300	0.9976	0.9108	0.2632	1.0000
400	1.0000	0.9832	0.4934	1.0000
500	1.0000	0.9992	0.7318	1.0000
1000	1.0000	1.0000	0.9990	1.0000

Table 10. Power Comparison for Different Normality Tests against Gamma (4, 5) alternative distribution at α = 0.05

	Simulated Power of test			
	α = 0.05			
n	**AD**	**CS**	**KS**	**SW**
10	0.1254	0.1178	0.0001	0.1334
20	0.2598	0.1330	0.0026	0.2976
30	0.3830	0.1812	0.0060	0.4554
40	0.4774	0.2290	0.0106	0.5780
50	0.5830	0.2652	0.0174	0.6788
100	0.8920	0.5160	0.0734	0.9568
200	0.9984	0.8630	0.3398	0.9998
300	1.0000	0.9752	0.6634	1.0000
400	1.0000	0.9966	0.8714	1.0000
500	1.0000	1.0000	0.9560	1.0000
1000	1.0000	1.0000	1.0000	1.0000

Table 11. Power Comparison for Different Normality Tests against Gamma (4, 5) alternative distribution at α = 0.10

	Simulated Power of test			
	α = 0.10			
n	**AD**	**CS**	**KS**	**SW**
10	0.2058	0.1698	0.0016	0.2118
20	0.3452	0.2474	0.0110	0.3996
30	0.4826	0.2968	0.0236	0.5598
40	0.6048	0.3450	0.0372	0.6936
50	0.7062	0.3956	0.0666	0.8040
100	0.9370	0.6466	0.1790	0.9800
200	0.9996	0.9228	0.5594	1.0000
300	1.0000	0.9904	0.8536	1.0000
400	1.0000	0.9986	0.9558	1.0000
500	1.0000	1.0000	0.9924	1.0000
1000	1.0000	1.0000	1.0000	1.0000

At α = 0.05 level of significance, the SW test requires a sample size slightly higher than 50 to yield 80% power value. The sample size requirement of the AD test is similar to that of the SW test. The CS test will require sample size n > 100 to be 80% powerful. The KS test still requires the highest sample size to attain a comparable pre-specified power.

3.2.2. Power for Discrete Alternative Non-normal Distributions

The following tables show that the power of the tests varies with the significance levels, sample sizes and three different discrete alternative distributions. The tables show the power for selected alternative distributions for α = 0.01, 0.05 and 0.10.

Tables 12, 13, and 14 below show that the CS test is the most powerful of the four test for all significance levels considered. At $\alpha = 0.01$ and n = 30, the power of the CS test

is even more than our desired 80%. To obtain a pre-specified 80% power, the AD test will require a sample size lower than what will be required by the SW test in the range of n > 60. The KS test still require the highest sample size of n > 100 to be 80% powerful. As α increases, the sample size requirements for a desired power value of each test decreases.

Table 12. Power Comparison for Different Normality Tests against Binomial (n, 0.5) alternative distribution at α = 0.01

	Simulated Power of test			
	α = 0.01			
n	**AD**	**CS**	**KS**	**SW**
10	0.0636	0.0976	0.0000	0.0494
20	0.1132	0.3738	0.0012	0.0616
30	0.2200	0.8356	0.0016	0.1032
40	0.4328	0.9904	0.0030	0.1654
50	0.6760	1.0000	0.0100	0.2490
100	1.0000	1.0000	0.2166	0.9496
200	1.0000	1.0000	0.9960	1.0000
300	1.0000	1.0000	1.0000	1.0000
400	1.0000	1.0000	1.0000	1.0000
500	1.0000	1.0000	1.0000	1.0000
1000	1.0000	1.0000	1.0000	1.0000

Table 13. Power Comparison for Different Normality Tests against Binomial (n, 0.5) alternative distribution at α = 0.05

	Simulated Power of test			
	α = 0.05			
n	**AD**	**CS**	**KS**	**SW**
10	0.1952	0.3276	0.0014	0.1640
20	0.3698	0.6896	0.0104	0.2624
30	0.6342	0.9646	0.0270	0.3840
40	0.8712	0.9998	0.0636	0.5542
50	0.9828	1.0000	0.1298	0.7596
100	1.0000	1.0000	0.7768	1.0000
200	1.0000	1.0000	1.0000	1.0000
300	1.0000	1.0000	1.0000	1.0000
400	1.0000	1.0000	1.0000	1.0000
500	1.0000	1.0000	1.0000	1.0000
1000	1.0000	1.0000	1.0000	1.0000

Table 14. Power Comparison for Different Normality Tests against Binomial (n, 0.5) alternative distribution at α = 0.10

	Simulated Power of test			
	α = 0.10			
n	**AD**	**CS**	**KS**	**SW**
10	0.3186	0.4176	0.0056	0.2682
20	0.5806	0.8318	0.0290	0.4322
30	0.8356	0.9842	0.0920	0.5938
40	0.9756	1.0000	0.1918	0.8056
50	0.9994	1.0000	0.3442	0.9224
100	1.0000	1.0000	0.9684	1.0000
200	1.0000	1.0000	1.0000	1.0000
300	1.0000	1.0000	1.0000	1.0000
400	1.0000	1.0000	1.0000	1.0000
500	1.0000	1.0000	1.0000	1.0000
1000	1.0000	1.0000	1.0000	1.0000

Table 15. Power Comparison for Different Normality Tests against Poisson (4) alternative distribution at α = 0.01

	Simulated Power of test			
	α = 0.01			
n	**AD**	**CS**	**KS**	**SW**
10	0.0428	0.0492	0.0084	0.0382
20	0.0882	0.1554	0.0242	0.0706
30	0.1566	0.4056	0.0610	0.1186
40	0.2702	0.8046	0.0922	0.1930
50	0.4150	0.9804	0.1382	0.2816
100	0.9966	1.0000	0.2014	0.8688
200	1.0000	1.0000	0.9312	1.0000
300	1.0000	1.0000	1.0000	1.0000
400	1.0000	1.0000	1.0000	1.0000
500	1.0000	1.0000	1.0000	1.0000
1000	1.0000	1.0000	1.0000	1.0000

Table 16. Power Comparison for Different Normality Tests against Poisson (4) alternative distribution at α = 0.05

	Simulated Power of test			
	α = 0.05			
n	**AD**	**CS**	**KS**	**SW**
10	0.1522	0.2064	0.0004	0.1438
20	0.2880	0.3728	0.0054	0.2378
30	0.4506	0.6874	0.0184	0.3518
40	0.6704	0.9630	0.0510	0.5186
50	0.8528	0.9996	0.0994	0.6706
100	1.0000	1.0000	0.6274	0.9990
200	1.0000	1.0000	0.9998	1.0000
300	1.0000	1.0000	1.0000	1.0000
400	1.0000	1.0000	1.0000	1.0000
500	1.0000	1.0000	1.0000	1.0000
1000	1.0000	1.0000	1.0000	1.0000

Table 17. Power Comparison for Different Normality Tests against Poisson (4) alternative distribution at α = 0.10

	Simulated Power of test			
	α = 0.10			
n	**AD**	**CS**	**KS**	**SW**
10	0.2610	0.2792	0.0038	0.2350
20	0.4446	0.5338	0.0208	0.3870
30	0.6740	0.8250	0.0558	0.5498
40	0.8618	0.9798	0.1386	0.7100
50	0.9666	0.9998	0.2410	0.8354
100	1.0000	1.0000	0.8516	1.0000
200	1.0000	1.0000	1.0000	1.0000
300	1.0000	1.0000	1.0000	1.0000
400	1.0000	1.0000	1.0000	1.0000
500	1.0000	1.0000	1.0000	1.0000
1000	1.0000	1.0000	1.0000	1.0000

Tables 15, 16 and 17 show a fall in the power value of the four tests when compared with the binomial alternative

distribution. We also observed that the CS test is still the most powerful across all sample sizes and significance levels considered amongst the four tests. The AD test has a comparable power with the CS test while the KS is the least powerful in this collection. The sample size required by the CS to attain a desired power under the discrete distributions considered is lower than what was required for the three continuous alternative distributions earlier simulated.

3.2.3. Power of Tests for Mixture Distributions

Here, we considered the following mixture of two normal distributions:

i. Unequal means and equal variance with equal probabilities of mixture denoted Mixture Normal 1 as $\{N_1 (-1, 4) + N_2 (1, 4), p_1 = p_2 = 0.5\}$
ii. Unequal means and equal variance with unequal probabilities of mixture denoted Mixture Normal 2 as $\{N_1 (-1, 4) + N_2 (1, 4), p_1 = 0.3 \ \& \ p_2 = 0.7\}$
iii. Unequal mean and unequal variance with equal probabilities of mixture denoted Mixture Normal 3 as $\{N_1 (-1, 2) + N_2 (1, 4), p_1 = p_2 = 0.5\}$
iv. Unequal mean and unequal variance with unequal probabilities of mixture denoted Mixture Normal 4 as $\{N_1 (-1, 2) + N_2 (1, 4), p_1 = 0.3 \ \& \ p_2 = 0.7\}$

Where p_1 & p_2 are the mixture probabilities for the two normal distributions respectively.

The ranking procedure was adopted to obtain a clearer picture of the performance of the different normality tests. The rank 1 was given to the test with the highest power while rank 4 was assigned to the test which has the lowest power. The ranks were then summed to obtain the grand total of the ranks. Since the lowest rank was given to the test with the highest power, therefore the test which has the lowest sum of rank will be chosen as the most powerful test in our collection in detecting departure from normality. The following tables show the rank of power based on the type of alternative distribution and sample sizes, respectively.

Table 18. Power Comparison against Mixture Normal 1 at α = 0.05

	Simulated Power of test			
	α = 0.05			
n	**AD**	**CS**	**KS**	**SW**
10	0.0448	0.0666	0.0484	0.0416
20	0.0476	0.0450	0.0478	0.0448
30	0.0502	0.0492	0.0536	0.0518
40	0.0478	0.0636	0.0492	0.0452
50	0.0450	0.0498	0.0528	0.0438
100	0.0502	0.0504	0.0482	0.0460
200	0.0464	0.0508	0.0458	0.0454
300	0.0514	0.0522	0.0466	0.0480
400	0.0552	0.0524	0.0500	0.0470
500	0.0488	0.0458	0.0446	0.0476
1000	0.0564	0.0568	0.0590	0.0514

Table 19. Power Comparison against Mixture Normal 2 at α = 0.05

	Simulated Power of test			
	α = 0.05			
n	**AD**	**CS**	**KS**	**SW**
10	0.0488	0.0656	0.0552	0.0474
20	0.0452	0.0504	0.0524	0.0534
30	0.0512	0.0486	0.0490	0.0520
40	0.0490	0.0556	0.0508	0.0494
50	0.0518	0.0538	0.0530	0.0500
100	0.0520	0.0536	0.0442	0.0532
200	0.0584	0.0602	0.0492	0.0534
300	0.0590	0.0546	0.0550	0.0626
400	0.0682	0.0524	0.0556	0.0686
500	0.0768	0.0566	0.0532	0.0726
1000	0.1030	0.0642	0.0588	0.1112

Table 20. Power Comparison against Mixture Normal 3 at α = 0.05

	Simulated Power of test			
	α = 0.05			
n	**AD**	**CS**	**KS**	**SW**
10	0.0694	0.0748	0.0520	0.0626
20	0.0920	0.0672	0.0586	0.0882
30	0.1104	0.0774	0.0672	0.1140
40	0.1330	0.0876	0.0732	0.1302
50	0.1572	0.0824	0.0816	0.1542
100	0.2812	0.1192	0.0894	0.2920
200	0.5246	0.2092	0.1446	0.5396
300	0.7204	0.2886	0.1900	0.7264
400	0.8296	0.3676	0.2400	0.8546
500	0.9116	0.4510	0.2994	0.9094
1000	0.9974	0.8030	0.5488	0.9986

Table 21. Power Comparison against Mixture Normal 4 at α = 0.05

	Simulated Power of test			
	α = 0.05			
n	**AD**	**CS**	**KS**	**SW**
10	0.0538	0.0684	0.0500	0.0516
20	0.0596	0.0528	0.0498	0.0604
30	0.0652	0.0612	0.0554	0.0624
40	0.0642	0.0664	0.0538	0.0678
50	0.0850	0.0584	0.0606	0.0822
100	0.1180	0.0718	0.0586	0.1220
200	0.2030	0.0984	0.0730	0.2176
300	0.3014	0.1134	0.0954	0.3160
400	0.4112	0.1514	0.1048	0.4158
500	0.4830	0.1704	0.1246	0.5062
1000	0.8118	0.3002	0.2136	0.8238

Table 22. Rank of Power of tests under all three continuous alternative distributions at α = 0.01

	α = 0.01			
n	**AD**	**CS**	**KS**	**SW**
10	8	5	8	9
20	6	7	9	8
30	7	8	9	6
40	6	10	9	5
50	5	10	11	4
100	6	9	12	3
200	6	9	12	3
300	5.5	9	12	3.5
400	5	9	12	4
500	5.5	8	12	4.5
1000	5.5	7	12	5.5
Sum of ranks	65.5	91.0	118.0	55.5

Table 23. Rank of Power of tests under all three continuous alternative distributions at α = 0.05

	α = 0.05			
n	**AD**	**CS**	**KS**	**SW**
10	7	6	9	8
20	6	11	8	5
30	6	11	10	3
40	6	10	11	3
50	6	10	11	3
100	6	9	12	3
200	5.5	9	12	3.5
300	5	9	12	4
400	5	9	12	4
500	6	7	12	5
1000	6	7.5	10.5	6
Sum of ranks	64.5	98.5	119.5	47.5

Table 24. Rank of Power of tests under all three continuous alternative distributions at α = 0.10

	α = 0.10			
n	**AD**	**CS**	**KS**	**SW**
10	8	9	8	5
20	8	8	9	5
30	6	10	11	3
40	6	10	11	3
50	6	10	11	3
100	6	9	12	3
200	5.5	9	12	3.5
300	5	9	12	4
400	5	9	12	4
500	5.5	7	12	5.5
1000	6.5	8	9	6.5
Sum of ranks	67.5	98	119	45.5

In summary, the sums ranks of power of the four tests under the continuous, discrete distributions and mixture normals are presented.

Table 25. Total Rank of Power based on the type of Alternative distribution

	α = 0.01			
Alternative Distributions	**AD**	**CS**	**KS**	**SW**
Continuous	65.5	91.0	118.0	55.5
Discrete	47.5	36.5	76.0	60.0
	α = 0.05			
Continuous	64.5	98.5	119.5	47.5
Discrete	48.0	38.0	74.5	59.5
Mixture 1	25.0	21.0	25.0	39.0
Mixture 2	28.0	26.0	33.0	23.0
Mixture 3	18.0	31.0	41.0	17.0
Mixture 4	21.0	31.0	43.0	15.0
	α = 0.10			
Continuous	67.5	98.0	119.0	45.5
Discrete	47.0	39.0	73.0	59.0

Power analysis show that the choice of a normality test should be made with special consideration for the type of measurement in which the observed data are collected. Under the three continuous alternative distribution considered in this study, Shapiro-Wilk test is the most powerful test amongst the four tests for all significance levels considered. The Anderson-Darling test may also be adopted in place of the Shapiro-Wilk test due to its reasonable power comparison against the Shapiro-Wilk test. Chi-square test and Kolmogorov-Smirnov test have low power for the continuous alternative distributions considered.

Chi-square test is the most powerful of the four tests under the two discrete alternative distributions considered. It consistently demonstrated the highest power for all significance levels considered. The Anderson-Darling test again is next to the Chi-square test, followed by Shapiro-Wilk test. The Kolmogorov-Smirnov test is the least powerful among the four normality tests considered in this work.

The power of all four tests, regardless of the two mixture probabilities considered, were very low under the mixture of two normal distributions with unequal means and equal variance. However, with a mixture probability of 0.5 each, Shapiro-Wilk test out-performed the other three tests under mixture of two normals with unequal means and unequal variances. This is closely followed by the Anderson-Darling test, then the Chi-Square and Kolmogorov Smirnov tests respectively.

We also found that the power of the four tests generally increase as the sample size and significance level increase as expected theoretically.

4. Conclusions

None of the four tests considered in this study is uniformly most powerful for all types of distributions, sample sizes and significance levels considered.

For continuous alternative distributions, Shapiro-Wilk test is the most powerful test for all sample sizes whereas Kolmogorov-Smirnov test is the least powerful test in our collection. However, the power of Shapiro-Wilk test is still low for small sample size. The performance of Anderson-Darling test is quite comparable with Shapiro-Wilk test.

For discrete alternative distributions, Chi-square test outperforms the other three tests at all sample sizes. Anderson-Darling test is next to it while Shapiro-Wilk test performs better than Kolmogorov-Smirnov test.

Given the inconsistencies in power performance of all four tests under mixture of two normal distributions, we recommend that more work should be carried out on the search for normality tests which will perform better under these special conditions and the effect of mixture probabilities on the performance of normality tests.

REFERENCES

[1] Althouse, L.A., Ware, W.B. and Ferron, J.M. (1998). Detecting Departures from Normality: A Monte Carlo Simulation of a New Omnibus Test based on Moments.

[2] Anderson, T.W. and Darling, D.A. (1954). A Test of Goodness of Fit. *Journal of the American Statistical Association*, Vol. 49, No. 268, 765-769.

[3] Arshad, M., Rasool, M.T. and Ahmad, M.I. (2003). Anderson Darling and Modified Anderson Darling Test for Generalized Pareto Distribution. *Pakistan Journal of Applied Sciences* 3(2), pp. 85-88.

[4] Conover, W.J. (1999). Practical Nonparametric Statistics. Third Edition, John Wiley & Sons, Inc. New York, pp. 428-433 (6.1).

[5] D'Agostino, R. and Pearson, E.S. (1973). Test for Departures from Normality. Empirical Results for the Distribution of b_2 *and* $\sqrt{b_1}$. *Biometrika*, Vol.60, No. 3, pp.613-622.

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

[7] Dufour J.M., Farhat, A., Gardiol, L. and Khalaf, L. (1998). Simulation-based Finite Sample Normality Tests in Linear Regression. *Econometrics Journal*, Vol.1, pp.154-173.

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

[9] Jarque, C.M. and Bera, A.K. (1987). A test for normality of observations and regression residuals, *Internat. Statst. Rev.* 55(2), pp.163-172.

[10] Kolmogorov, A.N. (1933). Sulla determinazioneempirica di unalegge di distribuzione, Giornaledell' Instituto Italianodegli Attuari 4, pp. 83-91 (6.1).

[11] Lilliefors, H.W. (1967). On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. *Journal of American Statistical Association.* Vol. 62, No. 318, pp. 399-402.

[12] Mendes, M. and Pala, A. (2003). Type I error Rate and Power of Three Normality Tests. *Pakistan Journal of Information and Technology* 2(2), pp. 135-139.

[13] Normadiah, M.R. and Yap, B.W. (2010). Power Comparisons of some selected normality tests. *Proceedings of the Regional Conference on Statistical Sciences*, pp. 126-138.

[14] Park, H.M. (2008). Univariate Analysis and Normality Test Using SAS, Stata, and SPSS. Technical Working Paper. The University Information Technology Services (UITS) Center for Statistical and Mathematical Computing, Indiana University.

[15] Royston, P. (1992). Approximating the Shapiro-Wilk W test for Normality [Abstract]. *Statistics and Computing*, 2, pp.117-119.

[16] Seier, E. (2002). Comparison of Tests for Univariate Normality. *InterStat Statistical Journal*, 1, pp.1-17.

[17] Shapiro, S.S. and Wilk, M.B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, Vol. 52, No. 3, pp. 591-611.

[18] Thadewald, T. and Buning, H. (2007). Jarque-Bera and its Competitors for Testing for Normality. *Journal of Applied Statistics*, Vol. 34, No.1, pp. 87-105.

[19] Thode, H.C. Testing for Normality. Marcel Dekker, New York, 2002.