British Journal of Applied Science & Technology
4(18): 2646-2660, 2014

SCIENCEDOMAIN ***international***
www.sciencedomain.org

Normality Assessment, a Substantial Appraisal in Medical Studies: A Simulation Study for Power Comparison of Various Types of Normality Tests

Hamed Tabesh[1], Ahmad Heidari[1] and Azadeh Saki[1*]

[1]Department of Biostatistics and Epidemiology, School of Health, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran.

Authors' contributions

This work was carried out in collaboration between all authors. Author HT designed the study, wrote the first draft of manuscript. Author AH performed the simulation, wrote the protocol and managed literature searches. Author AZ conceptualized and managed the study. All authors read and approved the final manuscript.

Original Research Article

Received 4th January 2014
Accepted 9th March 2014
Published 10th May 2014

ABSTRACT

Background: In medical research, statistical tests have become more important. Several parametric procedures are available but each of them requires normality assumption. As normality violation may affect interpretation and inferences reliability and validity, so importance of normal distribution is undeniable. Although several normality tests available for software users but the power of each test in specified situation is not clear.
Methods: The aim of this study is to compare the power of nine normality tests. In this paper power of Jarque-Bera test, D'Agostino and Pearson test, Chi-square test, Kolmogrov-Smirnov test, Lilliefors test, Cramer-Von Mises test, Anderson-Darling test, Shapiro-Wilk test and Shapiro-Francia test compared via Monte Carlo simulation of sample data generated from alternative distributions that follow symmetric, skewed, skewed & heavy tailed, highly skewed and highly skewed & heavy tailed distributions.
Results: Simulation study shows that Shapiro-Francia test under symmetric, skewed, skewed& heavy tailed and highly skewed& heavy tailed distributions perform better than others. Also, Shapiro-Wilk performs better when underlying distribution is skewed.
Conclusion: Although, Shapiro-Francia and Shapiro-Wilk have greater power than their

**Corresponding author: E-mail: azadehsaki@yahoo.com;*

competitors but their powers are still low for small sample size. So their singly use is not recommended.

Keywords: Normality test; Power; Monte Carlo Simulation; Tukey g-and-h distribution.

1. INTRODUCTION

Normality is one of the serious assumptions of many parametric tests and methods. Methods such as linear regression, analysis of variance and discriminant analysis are not robust with respect to abnormality [1-4]. Without normality, interpretations and inferences from tests with normality assumption become invalid [1,3]. For example when independent samples t-test used , if normality assumption is violated ,the reliability of the test may be compromised [5]. Generally, for parametric tests estimations based on prediction and modeling the data without passing normality assumption are not valid [1]. Normality assumption of the linear regression error distribution in most situations will create favorable conditions for analysis because the error represents the effect of all eliminated factors from the model. Also according to sensitivity of t-test for large distance form normality, when the errors are great especially when distribution is highly skewed, estimated confidence interval would not be accurate [6]. Parametric tests are valid when the normality assumption is not violated [7]. So, before applying parametric test, assessment of normality is essential [2]. In the medical studies, estimations, comparisons and statistical tests are of particular importance because findings of a research might be caused changing treatment in therapy program or might be influenced how to care patients. In the other words, researches in this area directly or indirectly will affect health, treatment and quality of life so, error in humans conclusion could be endangering more than any other sciences. Also, normality distribution for a variable could be a great achievement to find reference curves easily for that index. Several statistical procedures test normality in particular circumstances [4,8]. In most of the available software, some tests are intended to assess normality, e.g. statistical package for social sciences (SPSS) utilizes Kolmogrov-Smirnov test, Shapiro-Wilk test and Lilliefors test; Graphpad Prism software uses D'Agostino and Pearson test, Kolmogrov-Smirnov test and Shapiro-Wilk test; MINITAB software applies Anderson-Darling test, Ryan-Joiner test and Kolmogrov-Smirnov test; STATISTICA software utilizes Shapiro-Wilk test, Kolmogrov-Smirnov test, Lilliefors test and Chi-square test; and R program uses Shapiro-Wilk test, Shapiro-Francia test and some other tests. These are examples of regular normality tests in world of software, that show us the dependency of type of normality test which utilizes and software which uses [7].

Using SPSS and Graphpad Prism among users of medical field is more common, because of their ease of use. According to the studies in last decade, many of these tests have relatively good power for normality diagnostics when sample size is larger than 50 but when sample size is small the power lead to less than 60% that could not be acceptable [1,7].

The latter issue is not clear for users of these tests and users apply them with lack of knowledge about their power weakness. So, collection of normality tests for different conditions or a test that can be the best for all situations seems to be essential, although none of the currently popular software provides.

Different methods are available to check whether the distribution of observations conform to a normal distribution [9]. In general normality assumption assesses in two ways, graphical procedure and normality test [3,9].

Graphical procedures such as histogram and box plot, give information about data distribution [9]. Although they are useful scheme for normality assumption they could not ensure that the distribution is normal [7,9].

Decision making based on graphical procedure is subjective and it is probable that this decision differ from reality. So, in most cases confirmatory methods is required [7]. Normality tests are often classified into four categories as follows:

1- Moment ratio techniques
2- Goodness of fit tests
3- Tests based on empirical distribution function
4- Tests based on correlation [10].

For the first class, Jarque-Bara test and D'Agostino and Pearson test could be mentioned [1].

Also Chi-square test would be the most famous normality test of the second category [7]. Most regular tests in third category are Kolmogrov-Smirnov test, Lilliefors test, Cramer-Von Mises test and Anderson-Darling test [8]. Without any doubt Shapiro-Wilk test and Shapiro-Francia test are good examples for the last category [11].

Each of these aforementioned tests could be used by some of statistical software. In this study, we wish to calculate power of the most powerful, regular and recommended univariate normality tests in previous researches via simulation studies for different distributions (symmetric, skewed with light tailed, skewed with heavy tailed, highly skewed with light tailed and highly skewed with heavy tailed).

Let $X_1, X_2, \dots, X_n$ be a random sample of independently and identically distributed random samples from X, continuous univariate distribution then the formal testing whether the observed sample comes from a normally distribution population as follow:

$$\begin{cases} H_0: X \sim N(\mu\,, \sigma^2) \\ \quad H_1: not H_0 \end{cases} \tag{1}$$

A test is said to be powerful when it has a high probability of rejection H_0 when the sample come from a non-normal distribution.

This simulation study focuses on the performance of nine selected normality test from different categories of normality test via Monte Carlo simulation by computing rejection proportion when underlying distribution is not normal.

2. MATERIALS AND METHODS

A total of 9 different tests of normality with at least one published document which marked them as one of the three most powerful tests, were given in this study.

2.1 Test for Normality

In this section, for completeness a brief review of selected test statistics is presented.

2.1.1 Jarque–bera test

The Jarque–Bera test (JB) statistic is based on sample skewness $\sqrt{b_1}$ and kurtosis b_2 and is given as,

$$JB=n\left(\frac{(\sqrt{b_1})^2}{6}+\frac{(b_2-3)^2}{24}\right)$$

The JB statistic follows approximately a Chi-square distribution with two degrees of freedom [12].

2.1.2 D'Agostino and pearsontest

The D'Agostino and Pearson test (DP) statistic is

$$DP = Z^2(\sqrt{b_1}) + Z^2(b_2),$$

where $Z(\sqrt{b_1})$ and $Z(b_2)$, are the normal approximations to sample skewness $\sqrt{b_1}$ and kurtosis b_2 respectively. The DP statistic follows approximately a Chi-square distribution with two degrees of freedom [13].

2.1.3 Chi-square test

Chi-square test (CSQ) compares observed and expected (i.e., the hypothesized distribution) frequencies for individual categories, where n is the number of cells. The CSQ test statistic CSQ is given by,

$$CSQ = \sum_{i=1}^{n} \frac{(O_i-E_i)^2}{E_i},$$

If k parameters of the distribution of X need to be estimated, then distribution of CSQ follows approximately a CSQ distribution with n − k−1 degrees of freedom [14].

2.1.4 Kolmogrov-smirnov test

Kolmogrov-Smirnov test (KS) is based on the maximum vertical difference between the empirical distribution function (EDF) and the normal cumulative distribution curve (when the null hypothesis is that the EDF demonstrates normality). Let $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ be an ordered random sample and the distribution of X is F(x). The test statistic is defined by,

$$KS = \sup_x |F^*(x) - F_n(x)|$$

where $F^*(x)$ is theoretical cumulative distribution function of the normal distribution with known mean, μ and standard deviation, σ, and $F_n(x)$ is the empirical distribution function (EDF)of the data [15].

2.1.5 Lilliefors test

Lilliefors test (LF) is a modification of the Kolmogorov-Smirnov test. It is suitable when the unknown parameters of the null distribution must be estimated from the sample data.

The LF statistic is defined by

$$\text{LF} = max_x|F^*(X) - S_n(X)|$$

where $S_n(X)$ is the sample cumulative distribution function and $F^*(X)$ is the cumulative normal distribution function with $\mu = \bar{x}$, the sample mean and s^2, the sample variance [16,17].

2.1.6 Cramer-von mises test

The Cramer-von Mises test (CVM) statistic is defined by,

$$CVM = n\int_{-\infty}^{\infty}\{\mathrm{F_n(x)} - \mathrm{F(x)}\}^2[\mathrm{F(x)}]\mathrm{dF(x)}$$

Also it can be computed as, $CVM = \frac{1}{12n} + \sum_{i=1}^{n}\left[F_0(x_{(i)}) - \frac{2i-1}{2n}\right]^2$ [18].

2.1.7 Anderson-darling test

Anderson-Darling test (AD) is a modification of the Cramer-Von Mises test (CVM). It differs from the CVM in such a way that it gives more weight to the tails of the distribution. The AD test statistic is,

$$\text{AD}=\int_{-\infty}^{\infty}[\mathrm{F_n}(\mathrm{x}) - \mathrm{F(x)}]^2\psi(\mathrm{F(X)})\mathrm{dF(x)},$$

By taking ψ (F(x)) = 1, the AD statistic reduces to the CVM statistic.

This study used the following modified AD statistic given by, $\text{AD}^*=(1+\frac{0.75}{n} + \frac{2.25}{n^2})$ AD [19].

2.1.8 Shapiro-wilk test

The Shapiro-Wilk Test (SW) statistic is defined by $\text{SW}=\frac{(\sum_{i=1}^{n} a_i x_{(i)})^2}{\sum_{i=1}^{n}(x_{(i)}-\bar{x})^2}$,

Where $x_{(i)}$ is the i[th] order statistic,$\bar{x}$ is the sample mean,

$$a = (a_1, a_2, \ldots, a_n) = m'V^{-1}[(m'V^{-1})(V^{-1}m')]^{-\frac{1}{2}}$$

and $\boldsymbol{m} = (m_1,\cdots,m_n)'$ are the expected values of the order statistics of independent and identically distributed random variables sampled from the standard normal distribution and V is the covariance matrix of those order statistics [20,21].

2.1.9 Shapiro–francia test

The Shapiro–Francia test (SF) statistic is defined by,

$$\text{SF}=\frac{(\sum_{i=1}^{n} m_i x_{(i)})^2}{\sum_{i=1}^{n} m_i^2 \times \sum_{i=1}^{n}(x_{(i)}-\bar{x})^2},$$

Since SF equals the squared product-moment correlation coefficient between $x_{(i)}$ and m_i, small values of SF indicates non-normality [22-24].

2.2 Simulation

Monte Carlo simulation was used to compare the power of JB, DP, CSQ, KS, LF, CVM, AD, SW and SF test statistics in testing hypothesis (1), i.e. testing if a random sample of *n* independent observations come from a population with a normal distribution. Simulation were run for normal, non-normal symmetric, skewed, skewed & heavy tailed, highly skewed and highly skewed & heavy tailed distributions. All distributions except normal distribution generated using Tukey g-and-h distribution [25-27], i.e. generating Z_i from a normal distribution and setting X_i= (exp (gZ_i)-1)*exp ($hZ_i^2/2$))/g. For g=0 this expression is taken to be X_i= Z_i*exp ($hZ_i^2/2$) and for h=0, X_i= (exp (gZ_i)-1)/g.

As the g-and-h distribution provides a convenient method for considering a very wide range of situation corresponding to both symmetric and asymmetric distributions, it use is highly recommended. For Tukey g-and-h distribution g corresponds to skewness and h corresponds to kurtosis and tail heaviness [25]. The case g=0 corresponds to a symmetric distribution, and as g increases the skewness increases as well. In this study, simulation were run with g=0, 0.2, 0.5, 1 to span the range of skewness values that seems to occur in practice. Since skewness corresponds to g=k and g=-k are the same [25], so only positive values of g considered. The case h=0 corresponds to a distribution with similar kurtosis to standard normal distribution and as h increases the heaviness of tail increases as well. So simulations were run with h=0, 0.1, 0.2, 0.5. As the power of tests are depend on sample size, simulations were run for n= 8, 10, 15, 20, 25, 30, 40, 50, 100, 200. The Monte-Carlo study was employed, where 10,000 samples were generated for each combination of n=8, 10, 15, 20, 25, 30, 40, 50, 100, 200, g=0, 0.2, 0.5, 1 and h=0, 0.1, 0.2, 0.5.

3. RESULTS AND DISCUSSION

The power of the tests varies with sample size and alternative distributions. The results in Table1-10 were based on 10,000 samples of sizes 8, 10, 15, 20, 25, 30, 40, 50, 100, 200 respectively, alternative distributions as mentioned before generated by Tukey g-and-h distribution such as distributions with (g, h)=(0, 0.1), (0, 0.2),(0, 0.5) represent symmetric; (g, h)= (0.2, 0), (0.2, 0.1) represent skewed; (g, h)= (0.2, 0.2), (0.2, 0.5) represent skewed & heavy tailed; (g, h)= (0.5, 0), (0.5, 0.1), (1, 0), (1, 0.1) represent highly skewed and (g, h)= (0.5, 0.2), (0.5, 0.5), (1, 0.2), (1, 0.5) represent highly skewed & heavy tailed distributions. The sample sizes presented were sample sizes which frequently used in medical researches. It seems to be important to mention that DP test could not be used when sample size 20 or less [7].

Under almost all distributions for n=8, the simulation results, Table1, showed that SF test statistic performed better than any of the test statistics. It followed by SW and AD for symmetric, skewed and highly skewed distributions. Also SF followed by AD and SW for skewed & heavy tailed and highly skewed & heavy tailed distributions. However, for sample size n=8 powers of all tests for almost all distributions are less than 50%.

For sample size n=10 simulation results, Table2, showed similar patterns with slight changes in performance order between SW and AD. Power of any tests did not merit 60% for sample size n=10. Simulation results, Table 3, were closely similar for sample size n=15.

Under most of the alternative distributions for medium sample sizes,$20 \leq n \leq 50$, based on simulation results Tables 4,5,6,7,8 DP stated at the second rank after SF and they followed by SW and AD. Although as n increased power of test increased as well but for sample sizes n=20, 25, 30 power of tests did not further than 80% when underlying distributions were symmetric or skewed.

According to simulation results, Tables 9 and Table10, for large sample sizes, $n > 50$, SF test performed better than others and followed by JB and SW.

Table 1. Simulated power of normality tests for sample size 8 (n=8)

g, h	JB	CSQ	KS	LF	CVM	AD	SW	SF
0.0,0.0	0.0019	0.0302	0.0001	0.0526	0.0482	0.0524	0.0517	0.0573
0.0, 0.1	0.0092	0.0430	0.0001	0.0787	0.0841	0.0885	0.0893	0.1059
0.0, 0.2	0.0290	0.0734	0.0009	0.1242	0.1385	0.1428	0.1410	0.1653
0.0, 0.5	0.1327	0.2083	0.0214	0.3015	0.3266	0.3307	0.3220	0.3559
0.2, 0.0	0.0067	0.0419	0.0001	0.0674	0.0674	0.0716	0.0751	0.0828
0.2, 0.1	0.0168	0.0596	0.0002	0.0964	0.1054	0.1085	0.1108	0.1284
0.2, 0.2	0.0400	0.0869	0.0008	0.1452	0.1555	0.1622	0.1634	0.1874
0.2, 0.5	0.1398	0.2177	0.0236	0.3135	0.3359	0.3403	0.3328	0.3666
0.5 , 0.0	0.0358	0.1039	0.0004	0.1493	0.1696	0.1814	0.1880	0.2010
0.5, 0.1	0.0598	0.1294	0.0013	0.1814	0.2058	0.2156	0.2228	0.2368
0.5, 0.2	0.0842	0.1616	0.0045	0.2213	0.2440	0.2544	0.2578	0.2775
0.5, 0.5	0.1757	0.2726	0.0337	0.3655	0.3909	0.3966	0.3887	0.4191
1.0, 0.0	0.1504	0.3577	0.0103	0.3672	0.4386	0.4570	0.4786	0.4795
1.0, 0.1	0.1772	0.3595	0.0193	0.3880	0.4477	0.4637	0.4816	0.4830
1.0, 0.2	0.2021	0.3677	0.0307	0.4087	0.4601	0.4749	0.4852	0.4943
1.0, 0.5	0.2687	0.4156	0.0693	0.4833	0.5215	0.5297	0.5272	0.5491

Table 2. Simulated power of normality tests for sample size 10 (n=10)

g, h	JB	CSQ	KS	LF	CVM	AD	SW	SF
0.0,0.0	0.0098	0.0651	0.0000	0.0555	0.0507	0.0522	0.0512	0.0554
0.0, 0.1	0.0389	0.0926	0.0002	0.0908	0.0952	0.1011	0.1015	0.1187
0.0, 0.2	0.0883	0.1390	0.0023	0.1529	0.1645	0.1718	0.1715	0.1971
0.0, 0.5	0.2501	0.3290	0.0513	0.3707	0.4020	0.4058	0.3939	0.4397
0.2, 0.0	0.0222	0.0857	0.0000	0.0738	0.0792	0.0823	0.0864	0.0919
0.2, 0.1	0.0541	0.1144	0.0007	0.1107	0.1248	0.1312	0.1334	0.1541
0.2, 0.2	0.1039	0.1617	0.0040	0.1762	0.1936	0.2018	0.1997	0.2293
0.2, 0.5	0.2634	0.3450	0.0532	0.3786	0.4143	0.4218	0.4125	0.4509
0.5, 0.0	0.0924	0.1979	0.0015	0.1855	0.2177	0.2350	0.2478	0.2578
0.5, 0.1	0.1356	0.2229	0.0062	0.2245	0.2582	0.2729	0.2868	0.3009
0.5, 0.2	0.1802	0.2603	0.0159	0.2731	0.3102	0.3236	0.3287	0.3515
0.5, 0.5	0.3080	0.4049	0.0709	0.4403	0.4791	0.4835	0.4739	0.5121
1.0, 0.0	0.2909	0.5422	0.0256	0.4533	0.5537	0.5804	0.6051	0.5994
1.0, 0.1	0.3227	0.5201	0.0419	0.4768	0.5611	0.5786	0.5963	0.5985
1.0, 0.2	0.3501	0.5172	0.0641	0.5018	0.5708	0.5843	0.5939	0.6036
1.0, 0.5	0.4237	0.5700	0.1366	0.5840	0.6322	0.6405	0.6328	0.6548

Table 3. Simulated power of normality tests for sample size 15 (n=15)

g, h	JB	CSQ	KS	LF	CVM	AD	SW	SF
0.0,0.0	0.0201	0.0552	0.0001	0.0525	0.0531	0.0535	0.0529	0.0554
0.0, 0.1	0.0914	0.0801	0.0009	0.1024	0.1174	0.1254	0.1306	0.1583
0.0, 0.2	0.1892	0.1400	0.0086	0.1949	0.2283	0.2389	0.2431	0.2846
0.0, 0.5	0.4612	0.4104	0.1274	0.5034	0.5629	0.5709	0.5528	0.6045
0.2, 0.0	0.0536	0.0730	0.0003	0.0880	0.0960	0.1010	0.1099	0.1163
0.2, 0.1	0.1281	0.1058	0.0024	0.1404	0.1663	0.1744	0.1853	0.2099
0.2, 0.2	0.2199	0.1716	0.0166	0.2278	0.2655	0.2804	0.2855	0.3207
0.2, 0.5	0.4722	0.4286	0.1390	0.5165	0.5711	0.5800	0.5669	0.6150
0.5 , 0.0	0.2154	0.2200	0.0058	0.2617	0.3262	0.3568	0.3913	0.3876
0.5, 0.1	0.2892	0.2598	0.0225	0.3138	0.3817	0.4048	0.4282	0.4393
0.5, 0.2	0.3501	0.3132	0.0492	0.3804	0.4466	0.4623	0.4720	0.4926
0.5, 0.5	0.5370	0.5066	0.1818	0.5948	0.6494	0.6561	0.6403	0.6782
1.0, 0.0	0.5363	0.6874	0.0971	0.6548	0.7606	0.7906	0.8273	0.8083
1.0, 0.1	0.5773	0.6595	0.1460	0.6730	0.7586	0.7804	0.8048	0.7972
1.0, 0.2	0.6119	0.6469	0.1921	0.6903	0.7608	0.7757	0.7893	0.7928
1.0, 0.5	0.6822	0.6908	0.3248	0.7565	0.8049	0.8091	0.8023	0.8232

Table 4. Simulated power of normality tests for sample size 20 (n=20)

g, h	JB	DP	CSQ	KS	LF	CVM	AD	SW	SF
0.0,0.0	0.0213	0.0543	0.0493	0.0002	0.0480	0.0498	0.0491	0.0487	0.0516
0.0, 0.1	0.1324	0.1926	0.0786	0.0016	0.1069	0.1296	0.1408	0.1551	0.1846
0.0, 0.2	0.2715	0.3467	0.1557	0.0160	0.2272	0.2698	0.2897	0.3021	0.3505
0.0, 0.5	0.6056	0.6777	0.4926	0.2004	0.6037	0.6692	0.6778	0.6664	0.7124
0.2, 0.0	0.0822	0.1376	0.0744	0.0007	0.0984	0.1082	0.1200	0.1390	0.1397
0.2, 0.1	0.1875	0.2601	0.1090	0.0048	0.1659	0.1955	0.2122	0.2318	0.2548
0.2, 0.2	0.3126	0.3932	0.1861	0.0263	0.2751	0.3304	0.3518	0.3598	0.4041
0.2, 0.5	0.6212	0.6912	0.5140	0.2184	0.6219	0.6852	0.6938	0.6838	0.7308
0.5 , 0.0	0.3280	0.4316	0.2684	0.0115	0.3369	0.4309	0.4685	0.5198	0.5074
0.5, 0.1	0.4142	0.5097	0.3081	0.0445	0.4007	0.4870	0.5110	0.5416	0.5492
0.5, 0.2	0.4921	0.5747	0.3710	0.0914	0.4739	0.5525	0.5713	0.5839	0.6077
0.5, 0.5	0.6881	0.7485	0.6041	0.2936	0.7024	0.7557	0.7587	0.7495	0.7852
1.0, 0.0	0.7142	0.7974	0.8226	0.1936	0.7879	0.8790	0.9049	0.9316	0.9158
1.0, 0.1	0.7541	0.8225	0.7802	0.2614	0.8011	0.8760	0.8926	0.9115	0.9020
1.0, 0.2	0.7784	0.8361	0.7567	0.3276	0.8112	0.8733	0.8857	0.8952	0.8928
1.0, 0.5	0.8316	0.8706	0.8025	0.4951	0.8644	0.8997	0.9043	0.8990	0.9101

Table 5. Simulated power of normality tests for sample size 25 (n=25)

g, h	JB	DP	CSQ	KS	LF	CVM	AD	SW	SF
0.0,0.0	0.0259	0.0558	0.0584	0.0001	0.0479	0.0484	0.0469	0.0479	0.0494
0.0, 0.1	0.1690	0.2187	0.0934	0.0018	0.1143	0.1404	0.1564	0.1811	0.2144
0.0, 0.2	0.3427	0.4016	0.1958	0.0239	0.2601	0.3159	0.3368	0.3602	0.4118
0.0, 0.5	0.7154	0.7544	0.5980	0.2778	0.6874	0.7539	0.7627	0.7541	0.7965
0.2, 0.0	0.1068	0.1593	0.0888	0.0012	0.1083	0.1287	0.1433	0.1683	0.1658
0.2, 0.1	0.2429	0.3036	0.1405	0.0079	0.1885	0.2305	0.2527	0.2801	0.3093
0.2, 0.2	0.3943	0.4559	0.2405	0.0403	0.3216	0.3882	0.4100	0.4248	0.4720
0.2, 0.5	0.7291	0.7694	0.6180	0.2979	0.7057	0.7712	0.7807	0.7707	0.8120
0.5, 0.0	0.4264	0.5189	0.3555	0.0239	0.4141	0.5209	0.5654	0.6300	0.6092
0.5, 0.1	0.5269	0.6007	0.3875	0.0691	0.4874	0.5766	0.6023	0.6382	0.6410
0.5, 0.2	0.6033	0.6610	0.4588	0.1366	0.5634	0.6391	0.6573	0.6725	0.6935
0.5, 0.5	0.7920	0.8263	0.7079	0.3941	0.7869	0.8392	0.8457	0.8349	0.8637
1.0, 0.0	0.8375	0.8842	0.9076	0.3040	0.8751	0.9436	0.9603	0.9759	0.9671
1.0, 0.1	0.8643	0.9026	0.8647	0.3865	0.8834	0.9381	0.9497	0.9604	0.9548
1.0, 0.2	0.8801	0.9104	0.8490	0.4622	0.8895	0.9328	0.9413	0.9467	0.9477
1.0, 0.5	0.9131	0.9316	0.8875	0.6351	0.9231	0.9505	0.9506	0.9459	0.9547

Table 6. Simulated power of normality tests for sample size 30 (n=30)

g, h	JB	DP	CSQ	KS	LF	CVM	AD	SW	SF
0.0,0.0	0.0305	0.0557	0.0532	0.0001	0.0485	0.0498	0.0483	0.0483	0.0499
0.0, 0.1	0.2074	0.2495	0.0875	0.0027	0.1235	0.1585	0.1768	0.2066	0.2473
0.0, 0.2	0.4192	0.4584	0.2044	0.0316	0.2943	0.3660	0.3902	0.4197	0.4772
0.0, 0.5	0.7967	0.8162	0.6553	0.3539	0.7594	0.8216	0.8304	0.8207	0.8574
0.2, 0.0	0.1340	0.1871	0.0856	0.0011	0.1203	0.1472	0.1635	0.1966	0.1938
0.2, 0.1	0.2984	0.3496	0.1414	0.0104	0.2115	0.2659	0.2879	0.3219	0.3549
0.2, 0.2	0.4684	0.5128	0.2579	0.0539	0.3686	0.4447	0.4688	0.4873	0.5380
0.2, 0.5	0.8065	0.8271	0.6725	0.3772	0.7762	0.8360	0.8450	0.8337	0.8693
0.5, 0.0	0.5144	0.5924	0.3914	0.0362	0.4814	0.6020	0.6473	0.7179	0.6918
0.5, 0.1	0.6131	0.6739	0.4234	0.1002	0.5587	0.6491	0.6769	0.7131	0.7172
0.5, 0.2	0.6878	0.7336	0.5041	0.1861	0.6338	0.7138	0.7290	0.7449	0.7670
0.5, 0.5	0.8603	0.8764	0.7653	0.4835	0.8496	0.8935	0.8977	0.8886	0.9125
1.0, 0.0	0.9121	0.9349	0.9478	0.4270	0.9275	0.9728	0.9823	0.9921	0.9881
1.0, 0.1	0.9269	0.9477	0.9082	0.5129	0.9313	0.9669	0.9737	0.9821	0.9794
1.0, 0.2	0.9364	0.9523	0.8945	0.5817	0.9354	0.9636	0.9690	0.9721	0.9730
1.0, 0.5	0.9517	0.9615	0.9265	0.7410	0.9570	0.9748	0.9759	0.9721	0.9779

Table 7. Simulated power of normality tests for sample size 40 (n=40)

g, h	JB	DP	CSQ	KS	LF	CVM	AD	SW	SF
0.0,0.0	0.0330	0.0568	0.0569	0.0001	0.0516	0.0486	0.0485	0.0460	0.0489
0.0, 0.1	0.2667	0.2908	0.1054	0.0030	0.1390	0.1833	0.2053	0.2444	0.2972
0.0, 0.2	0.5271	0.5474	0.2572	0.0505	0.3600	0.4440	0.4741	0.5090	0.5742
0.0, 0.5	0.8934	0.8952	0.7744	0.4889	0.8607	0.9076	0.9136	0.9062	0.9307
0.2, 0.0	0.1854	0.2339	0.1044	0.0018	0.1444	0.1777	0.2005	0.2504	0.2476
0.2, 0.1	0.3936	0.4269	0.1761	0.0174	0.2590	0.3275	0.3567	0.4053	0.4422
0.2, 0.2	0.5904	0.6112	0.3318	0.0832	0.4487	0.5414	0.5666	0.5902	0.6390
0.2, 0.5	0.9043	0.9073	0.7931	0.5167	0.8751	0.9173	0.9227	0.9168	0.9367
0.5, 0.0	0.6739	0.7191	0.5098	0.0677	0.5922	0.7255	0.7778	0.8425	0.8205
0.5, 0.1	0.7515	0.7857	0.5343	0.1640	0.6692	0.7616	0.7891	0.8213	0.8255
0.5, 0.2	0.8107	0.8354	0.6211	0.2876	0.7469	0.8157	0.8339	0.8425	0.8580
0.5, 0.5	0.9386	0.9410	0.8706	0.6439	0.9237	0.9523	0.9555	0.9523	0.9654
1.0, 0.0	0.9800	0.9842	0.9883	0.6384	0.9806	0.9960	0.9984	0.9994	0.9991
1.0, 0.1	0.9843	0.9867	0.9709	0.7062	0.9796	0.9935	0.9959	0.9977	0.9972
1.0, 0.2	0.9839	0.9864	0.9597	0.7632	0.9813	0.9924	0.9937	0.9948	0.9948
1.0, 0.5	0.9892	0.9903	0.9746	0.8749	0.9894	0.9948	0.9948	0.9941	0.9954

Table 8. Simulated power of normality tests for sample size 50 (n=50)

g, h	JB	DP	CSQ	KS	LF	CVM	AD	SW	SF
0.0,0.0	0.0344	0.0545	0.0524	0.0001	0.0482	0.0501	0.0493	0.0491	0.0488
0.0, 0.1	0.3186	0.3306	0.0977	0.0044	0.1588	0.2097	0.2365	0.2834	0.3395
0.0, 0.2	0.6123	0.6132	0.2807	0.0661	0.4281	0.5198	0.5492	0.5835	0.6447
0.0, 0.5	0.9448	0.9420	0.8382	0.6019	0.9168	0.9519	0.9544	0.9511	0.9646
0.2, 0.0	0.2314	0.2757	0.0994	0.0025	0.1702	0.2059	0.2337	0.2974	0.2940
0.2, 0.1	0.4733	0.4951	0.1778	0.0244	0.3007	0.3802	0.4154	0.4687	0.5069
0.2, 0.2	0.6785	0.6866	0.3586	0.1154	0.5238	0.6184	0.6466	0.6710	0.7156
0.2, 0.5	0.9512	0.9448	0.8541	0.6308	0.9281	0.9569	0.9613	0.9563	0.9687
0.5 , 0.0	0.7922	0.8183	0.5831	0.1099	0.6996	0.8207	0.8649	0.9199	0.9023
0.5, 0.1	0.8467	0.8667	0.6018	0.2373	0.7601	0.8461	0.8684	0.8939	0.8936
0.5, 0.2	0.8870	0.8973	0.6899	0.3867	0.8256	0.8894	0.9010	0.9067	0.9191
0.5, 0.5	0.9727	0.9714	0.9200	0.7555	0.9663	0.9831	0.9835	0.9793	0.9856
1.0, 0.0	0.9962	0.9972	0.9971	0.7965	0.9950	0.9995	0.9999	1.0000	1.0000
1.0, 0.1	0.9965	0.9970	0.9882	0.8376	0.9949	0.9988	0.9997	0.9999	0.9996
1.0, 0.2	0.9954	0.9964	0.9822	0.8789	0.9943	0.9984	0.9991	0.9989	0.9994
1.0, 0.5	0.9967	0.9969	0.9905	0.9461	0.9968	0.9992	0.9997	0.9986	0.9991

Table 9. Simulated power of normality tests for sample size 100 (n=100)

g, h	JB	DP	CSQ	KS	LF	CVM	AD	SW	SF
0.0,0.0	0.0407	0.0531	0.0500	0.0002	0.0493	0.0503	0.0485	0.0495	0.0523
0.0, 0.1	0.5307	0.5024	0.1303	0.0079	0.2431	0.3265	0.3690	0.4562	0.5279
0.0, 0.2	0.8583	0.8310	0.4406	0.1544	0.6749	0.7790	0.8046	0.8317	0.8699
0.0, 0.5	0.9976	0.9966	0.9812	0.9082	0.9963	0.9988	0.9989	0.9984	0.9990
0.2, 0.0	0.4659	0.4926	0.1558	0.0070	0.3073	0.3818	0.4336	0.5468	0.5315
0.2, 0.1	0.7426	0.7397	0.2863	0.0726	0.5221	0.6179	0.6586	0.7183	0.7537
0.2, 0.2	0.9124	0.9007	0.5848	0.2905	0.7830	0.8635	0.8835	0.9025	0.9253
0.2, 0.5	0.9984	0.9978	0.9853	0.9260	0.9981	0.9997	0.9996	0.9993	0.9994
0.5 , 0.0	0.9893	0.9895	0.8877	0.3978	0.9467	0.9850	0.9932	0.9981	0.9966
0.5, 0.1	0.9883	0.9895	0.8696	0.5999	0.9634	0.9853	0.9895	0.9917	0.9923
0.5, 0.2	0.9926	0.9928	0.9179	0.7624	0.9812	0.9920	0.9932	0.9939	0.9958
0.5, 0.5	0.9998	0.9996	0.9970	0.9750	0.9997	1.0000	1.0000	1.0000	1.0000
1.0, 0.0	1.0000	1.0000	1.0000	0.9954	1.0000	1.0000	1.0000	1.0000	1.0000
1.0, 0.1	1.0000	1.0000	1.0000	0.9962	1.0000	1.0000	1.0000	1.0000	1.0000
1.0, 0.2	1.0000	1.0000	1.0000	0.9980	1.0000	1.0000	1.0000	1.0000	1.0000
1.0, 0.5	1.0000	1.0000	1.0000	0.9996	1.0000	1.0000	1.0000	1.0000	1.0000

Table 10. Simulated power of normality tests for sample size 200 (n=200)

g, h	JB	DP	CSQ	KS	LF	CVM	AD	SW	SF
0.0,0.0	0.0411	0.0501	0.0518	0.0001	0.0477	0.0486	0.0487	0.0491	0.0485
0.0,0.1	0.7673	0.7231	0.1904	0.0192	0.4072	0.5393	0.5966	0.6953	0.7561
0.0,0.2	0.9839	0.9737	0.7145	0.4124	0.9091	0.9611	0.9713	0.9779	0.9859
0.0,0.5	1.0000	1.0000	0.9998	0.9990	1.0000	1.0000	1.0000	1.0000	1.0000
0.2,0.0	0.7958	0.8050	0.2875	0.0385	0.5454	0.6717	0.7380	0.8465	0.8314
0.2,0.1	0.9392	0.9344	0.4900	0.2290	0.7944	0.8764	0.9004	0.9313	0.9426
0.2,0.2	0.9939	0.9916	0.8505	0.6488	0.9676	0.9881	0.9913	0.9930	0.9957
0.2,0.5	1.0000	1.0000	0.9999	0.9990	1.0000	1.0000	1.0000	1.0000	1.0000
0.5,0.0	1.0000	1.0000	0.9977	0.8827	0.9996	1.0000	1.0000	1.0000	1.0000
0.5,0.1	1.0000	1.0000	0.9952	0.9529	0.9996	0.9999	1.0000	1.0000	1.0000
0.5,0.2	1.0000	1.0000	0.9985	0.9873	0.9999	1.0000	1.0000	1.0000	1.0000
0.5,0.5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1.0,0.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1.0,0.1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1.0,0.2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1.0,0.5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Among considered tests in this study, SF, SW, AD, DP and JB had greater powers in different situations. So plots of these five tests are given in Fig.1 for non-normal symmetric distributions. Similarly; Figs. 2, 3, 4 and 5 showed rejection proportion plots of aforementioned five tests for skewed, skewed & heavy tailed, highly skewed and highly skewed & heavy tailed distributions respectively.

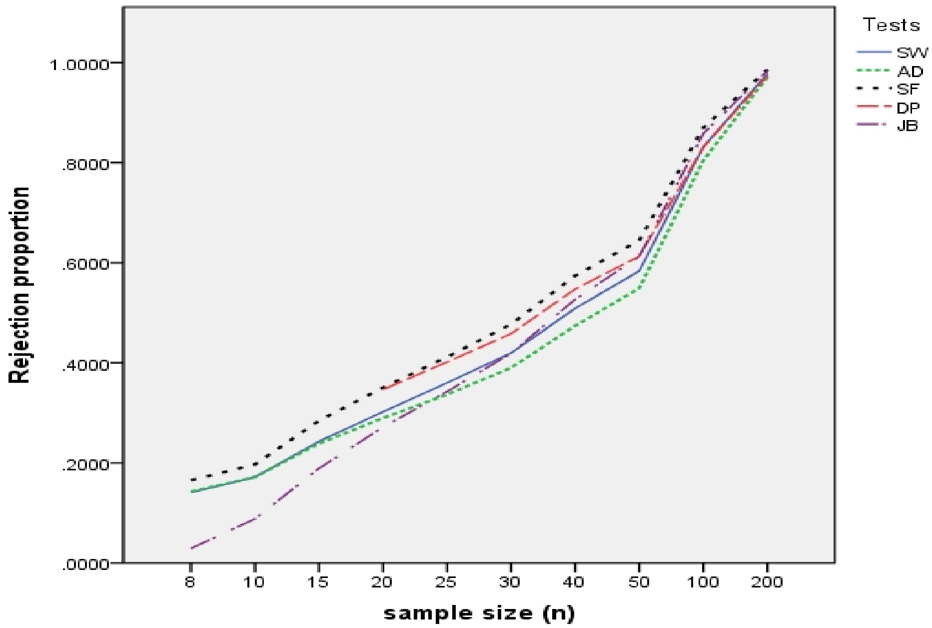

Fig. 1. Power comparison for preferable normality tests when underlying population is symmetric (g=0, h=.2)

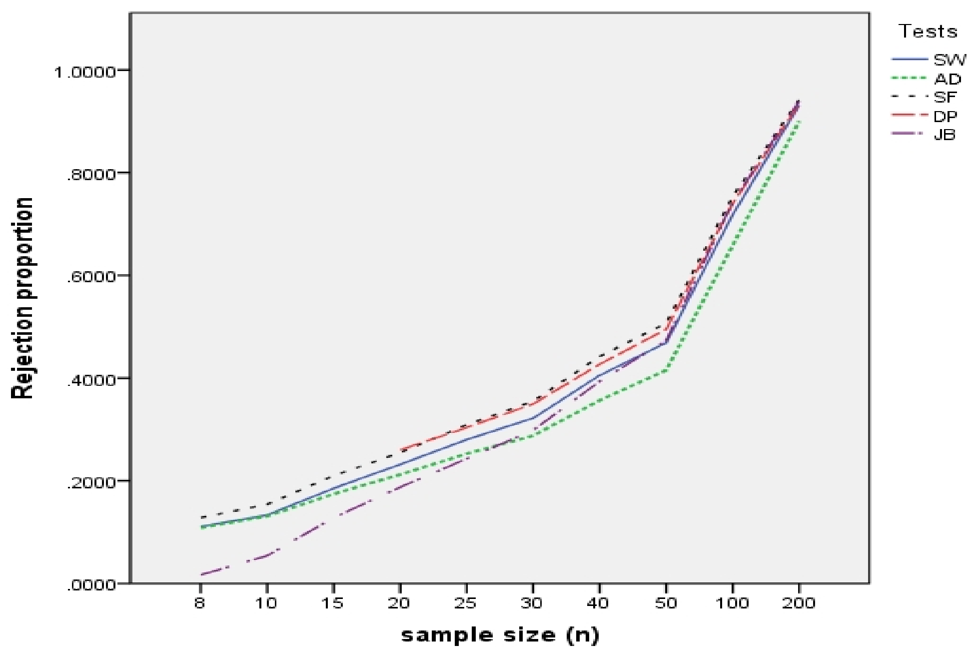

Fig. 2. Power comparison for preferable normality tests when underlying population is skewed (g=.2, h=.1)

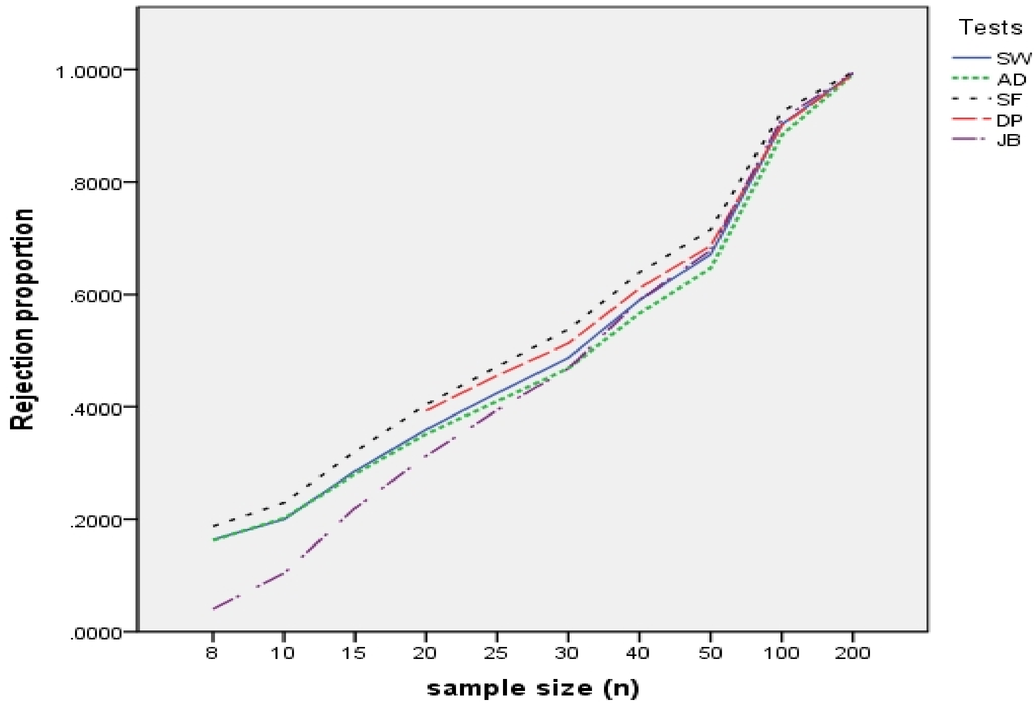

Fig. 3. Power comparison for preferable normality tests when underlying population is skewed & heavy tailed (g=.2, h=.2)

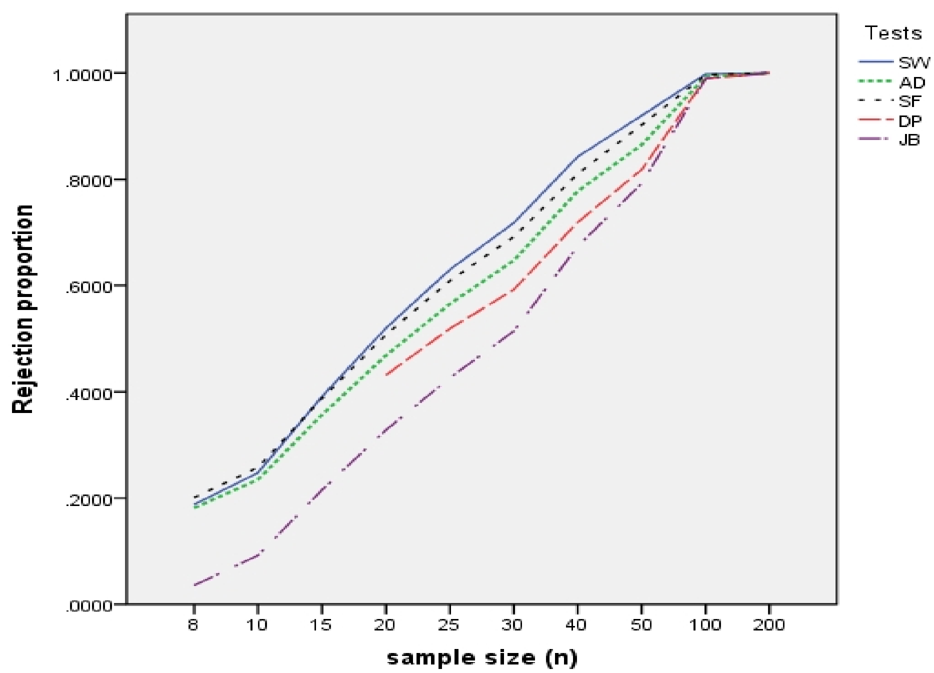

Fig. 4. Power comparison for preferable normality tests when underlying population is highly skewed (g=.5, h=0)

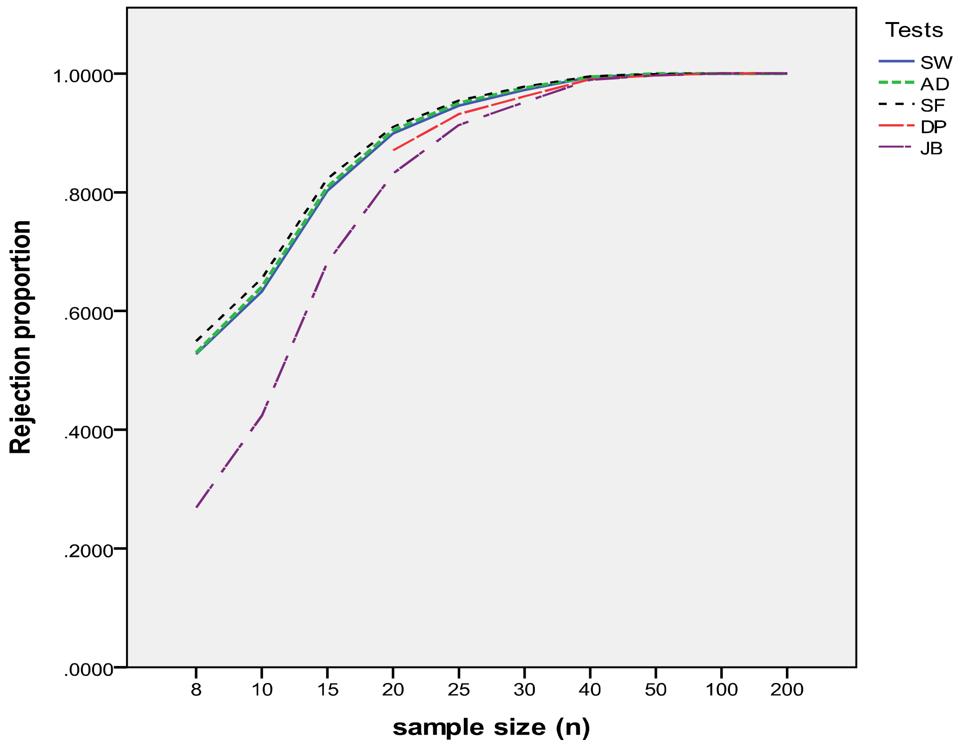

Fig. 5. Power comparison for preferable normality tests when underlying population is highly skewed & heavy tailed (g=1, h=.5)

4. CONCLUSION

Although nowadays statistical tests play an important role in medical research, many publications have been reported to contain serious statistical errors [28]. In this regard, violation of distributional assumption such as normality test has been identified as one of the most common problem. So make a decision about normality considered an important problem. According to number of normality available tests, we have compared them under various distributions with different sample sizes.

The results of the simulation studies showed that the Shapiro-Francia test performed better than most of its competitors whether the underlying distribution was normal, non-normal symmetric, skewed, skewed & heavy tailed or highly skewed & heavy tailed but not for highly skewed when $n \geq 15$. The simulation results also revealed that under highly skewed distribution, Shapiro-Wilk is the most powerful, although similar power could be considered for Shapiro-Wilk and Jarque-Bera for large sample sizes i.e. $n = 100{,}200$. Therefore, Shapiro-Francia's application where underlying distribution is not highly skewed recommended, since it is more powerful than any of alternatives compared here for almost all sample sizes. Also, where underlying distribution is highly skewed and there is small sample size. Shapiro- Wilk application's recommended where Shapiro-Francia is not the best.

Since these two tests with greater powers are not available in regular software that are used by researchers in medical field so using Kolmogrov-Smirnov test which is available in SPSS, MEDCALC and other friendly users software is not strange.

According to small power of all normality tests showed by simulation study in this research, the next challenge in this direction would be the preparation of a normality test for small sample sizes with greater powers that could be done with regular softwares.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Razali NM, Wah YB. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. Journal of Statistical Modeling and Analytics. 2011;2(1):21-33.
2. Laha A. A Note on Tests of Normality. Case and Teaching Notes, Forthcoming; 2006.
3. Park HM. Univariate analysis and normality test using SAS, STATA, and SPSS. The University Information Technology Services (UITS) Center for Statistical and Mathematical Computing, Indiana University; 2008.
4. Coin D. A goodness-of-fit test for normality based on polynomial regression. Computational statistics & data analysis. 2008;52(4):2185-2198.
5. Sheskin DJ. Parametric and nonparametric statistical procedures. 2th ed. Boca Raton: CRC; 2000.
6. Neter J, Wasserman W, Kutner MH. Applied linear statistical models. Vol. 4; 1996. Irwin Chicago.
7. Yap BW, Sim CH. Comparisons of various types of normality tests. Journal of Statistical Computation and Simulation. 2011;81(12):2141-2155.
8. Seier E. Comparison of tests for univariate normality. Interstat. 2002;1:1-17.
9. Öztuna D, Elhan AH, Tüccar E. Investigation of four Different normality tests in terms of type 1 error rate and power under different distributions. Turk J Med Sci. 2006;36(3):171-176.
10. Henderson AR. Testing experimental data for univariate normality. Clinica Chimica Acta. 2006;366(1):112-129.
11. Jarque CM, Bera AK. A test for normality of observations and regression residuals. International Statistical Review/Revue Internationale de Statistique. 1987;163-172.
12. D'Agostino RB, Belanger A,. D'Agostino RB Jr. A suggestion forusing powerful and informative tests of normality. The American Statistician. 1990;44(4):316-321.
13. Pearson ES, Bowman KO. Tests for departure from normality: Comparison of powers. Biometrika. 1977;64(2):231-246.
14. de Finetti B. Sulfapprossimazionc empirica di una legge di probabilità; 1933.
15. Lilliefors HW. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. Journal of the American Statistical Association. 1967;62(318):399-402.
16. Lilliefors HW. On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown. Journal of the American Statistical Association. 1969;64(325):387-389.
17. Cramér H. On the composition of elementary errors: First paper: Mathematical deductions. Scandinavian Actuarial Journal. 1928;1:13-74.

18. Anderson TW, Darling DA. A test of goodness of fit. Journal of the American Statistical Association. 1954;49(268):765-769.
19. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika. 1965;52(3/4):591-611.
20. Shapiro SS, Wilk MB, Chen HJ. A comparative study of various tests for normality. Journal of the American Statistical Association. 1968;63(324):1343-1372.
21. Shapiro SS, Francia RS. An approximate analysis of variance test for normality. Journal of the American Statistical Association. 1972;67(337):215-216.
22. Royston P. A toolkit for testing for non-normality in complete and censored samples. The statistician. 1993;37-43.
23. Royston P. A pocket-calculator algorithm for the shapiro-francia test for non-normality: An application to medicine. Statistics in medicine. 1993;12(2):181-184.
24. Jorge M, Boris I. Some propertiesof the Tukey g and h family of distributions. Communications in Statistics-Theory and Methods. 1984;13(3):353-369.
25. Field C, Genton MG. The multivariate g-and-h distribution. Technometrics. 2006;48(1).
26. Headrick TC, Kowalchuk RK, Sheng Y. Parametric Probability Densities and Distribution Functions for Tukey g-and-h Transformations and their Use for Fitting Data; 2008.
27. Rochon J. Gondan M, Kieser M. To test or not to test: Preliminary assessment of normality when comparing two independent samples. BMC medical research methodology. 2012;12(1):81.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sciencedomain.org/review-history.php?iid=510&id=5&aid=4523