



Empirical Type 1 Error Rate and Power Comparisons of Normality Tests with R

Ahmet PEKGOR, Murat ERISOGLU, Aydın KARAKOCA, Ulku ERISOGLU*

Necmettin Erbakan University, Department of Statistics, Konya, Turkey

Received: 23.05.2018; Accepted: 03.09.2018

<http://dx.doi.org/10.17776/csj.426382>

Abstract. Normality is one of the main presuppositions in statistical tests. The multiplicity of the normality tests bring out another problem of choosing the appropriate test for researchers. The free software R which has a great popularity in the statistical analysis has 18 normality tests in 4 different packages. In this study we compared performance of these normality tests in terms of empirical type 1 error rate and power by Monte Carlo simulation. As a result, regardless of the distribution of data (symetric or asymmetric) the Shapiro-Francia test, also the Frosini B test performed better than the other normality tests in terms of experimental type 1 error rate. However the widely used Kolmogorov-Smirnov test showed worse performance than other normality tests in terms of empirical type 1 error rate and power.

Keywords: Normality tests, R, Empirical type 1 error rate, Power of test.

R ile Normallik Testlerinin Deneysel 1. Tip Hata Oranı ve Güç Karşılaştırması

Özet. Normallik istatistiksel testlerde ana varsayımlardan biridir. Normallik testlerinin çokluğu, araştırmacılar için uygun normallik testini seçme konusunda başka bir problem ortaya çıkarmaktadır. İstatistiksel analizlerde sıklıkla kullanılan ücretsiz bir yazılım olan R’da 4 farklı pakette 18 normallik testi bulunmaktadır. Bu çalışmada, normallik testlerinin performansını ampirik 1. tip hata oranı ve güç açısından Monte Carlo simülasyonu ile karşılaştırılmıştır. Çalışma sonucunda verinin dağılımı (simetrik veya asimetrik) ne olursa olsun Shapiro-Francia testi, ayrıca Frosini B testi de deneysel tip 1 hata oranı açısından diğer normallik testlerinden daha iyi performans göstermiştir. Ayrıca yaygın olarak kullanılan Kolmogorov-Smirnov testi ampirik tip 1 hata oranı ve testin gücü açısından diğer normallik testlerinden daha kötü performans göstermiştir.

Anahtar Kelimeler: Normallik testi, R, Deneysel 1. tip hata oranı, Testin gücü

1. INTRODUCTION

Normality tests have a special importance in itself on statistical studies. Most of the parametric statistical tests have been developed under normality assumption such as one way analysis of variance, paired sample t test, Pearson correlation test etc. According to whether or not to met the normality assumption statistical inferences can give different results, and also ignoring the normality assumption can lead to great risks on the study in terms of methodically approach. Many normality tests have been developed since normality is important both in theoretical and applied studies. Initial work on normality tests began with Pearson’s [1] chi-square test early 20th century, continued with Cramer von Mises [2], Kolmogorov-Smirnov [3] and Geary [4]. Afterwards, a lot of normality tests have been developed by several researchers including Anderson Darling [5], Kuiper [6], Shapiro and Wilk [7], Lilliefors [8], Shapiro and Francia [9], Hegazy and Green [10], Weisberg and Bingham [11] Spiegelhalter [12], Frosini [13], Jargue and Bera [14] and Urzua’s corrected Jargue-Bera [15].

* Corresponding author. Email address: akarakoca@konya.edu.tr
<http://dergipark.gov.tr/csj> ©2016 Faculty of Science, Cumhuriyet University

In the following years, comparative studies among the numerous normality tests in the literature have come to the forefront. Esteban et. al. [16] compared four different normality tests based on entropy. Yazici and Yolacan [17] compared twelve different normality tests in terms of Powers by Monte-Carlo simulation. Yap and Sim [18] compared eight normality tests according to test power via generated data from symmetric and asymmetric distributions. Noughabi and Arghami [19] are investigated performance of the seven normality tests for different situations.

In this study, we aim to help researchers for choosing the appropriate normality test. R is one of the most popular statistical freeware and has 18 normality tests. In this study, we compared the empirical type 1 error rate and power of these normality tests. The power comparisons of tests were conducted according to different sample size for symmetric and asymmetric distributions.

2. TEST OF NORMALITY

Suppose X_1, X_2, \dots, X_n is a random sample from a univariate continuous distribution which has unknown probability density function (pdf) and parameter vector respectively $f(x, \Theta)$ and $\Theta = (\theta_1, \theta_2, \dots, \theta_p)'$. Testing of whether the sample comes from the normal distribution can be tested by the hypothesis, $H_0: f(x, \Theta) \in N(\mu, \sigma^2)$ against $H_1: f(x, \Theta) \notin N(\mu, \sigma^2)$.

Normality tests can be grouped as regression and correlation based tests, empirical distribution tests, moment tests and other special tests [18]. Shapiro-Wilk (SW), Shapiro-Francia (SF) and Weisberg-Bingham (WB) tests are normality tests based on regression and correlation. These tests based on correlation coefficient are computed between order statistics in the sample and scores which denoted what the order statistics should be if the population is normal.

Let $X_{(i)}$ be the i th largest order statistic and \bar{X} be the sample mean. The SW test statistic for normality is given by

$$SW = \frac{(\sum_{i=1}^n a_i X_{(i)})^2}{\sum_{i=1}^n (X_{(i)} - \bar{X})^2}$$

where a_i 's are coefficients and these are found from created table by Royston [20]. The SF test statistic has the similar form as the SW test as below,

$$SF = \frac{(\sum_{i=1}^n m_i X_{(i)})^2}{\sum_{i=1}^n (X_{(i)} - \bar{X})^2}$$

where m_i represents a vector of standard normal ordered statistics.

Another test statistic based on correlation is proposed by Weisberg and Bingham [11]. The WB test statistic is defined by

$$WB = \frac{(\sum_{i=1}^n c_i X_{(i)})^2 / \sum_{i=1}^n c_i^2}{\sum_{i=1}^n (X_{(i)} - \bar{X})^2}$$

where $c_i = \Phi^{-1}\left(\frac{i-3/8}{n+1/4}\right)$.

Kolmogorov-Smirnov (KS), Lilliefors (LF), Cramer Von Mises (CM), Anderson-Darling (AD) and Frosini's (B) tests are improved based on the difference between empirical and Normal distributions. The empirical distribution function (EDF) is the cumulative distribution function associated with the empirical measure of a sample. The empirical distribution function $S_n(\cdot)$ is defined as

$$S_n(x) = \begin{cases} 0, & x < X_{(1)} \\ \frac{i}{n}, & X_{(i)} \leq x < X_{(i+1)}, \quad i \in \{1, \dots, n-1\} \\ 1, & X_{(n)} \leq x \end{cases}$$

where $X_{(i)}$ is the order statistic. The KS test statistic relates to supremum class of EDF statistics and is based on the largest vertical difference between the hypothesized Normal $F_0(X_{(i)})$ and empirical distributions. The KS test statistic is defined as

$$KS = \max(D^+, D^-)$$

where D^+ and D^- are $D^+ = \sup[[F_0(X_{(i)}) - S_n(X_{(i)})]]$ and $D^- = \sup[[F_0(X_{(i)}) - S_n(X_{(i-1)})]]$ respectively. The larger values of KS indicate non-normality.

The LF test statistic is a modification of the KS test. This test compares the empirical distribution with the obtained normal distribution which estimated from sample data.

The CM statistic is an omnibus test and computed by following equation.

$$CM = \frac{1}{12} + \sum_{i=1}^n \left[F_0(X_i) - \frac{2i-1}{2n} \right]^2$$

The AD test statistic is a modification of the CM test statistic. The AD test gives more weight to the tails than does the CM test. This test statistic is given by

$$AD = -n - \frac{1}{n} \sum (2i-1)[\log S_n(X_i) + \log(1 - S_n(X_{n+1-i}))].$$

The Frosini's B test statistic is the absolute sum of all differences between empirical distribution defined as $\frac{i-0.5}{n}$ and the standard normal distribution CDF of the i th value. The B test statistics is defined mathematically as

$$B = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left| \Phi(z_i) - \frac{i-0.5}{n} \right|.$$

The skewness (S), the kurtosis (K), Jarque Bera (JB), D'Agostino (DA) and Adjusted Jargue-Bera (AJB) tests are normality tests based on moments. Moments tests are those obtained from the identification that the departure from normality may be detected based on skewness and kurtosis [21]. For the normal distribution, the theoretical value of skewness is zero, and the theoretical value of kurtosis is three.

The sample estimates of skewness and kurtosis coefficients are computed as respectively

$$\alpha_3 = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^3}{\left\{ \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right\}^{3/2}}$$

$$\alpha_4 = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^4}{\left\{ \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right\}^2}$$

the test statistic of the skewness (S) test is denoted by

$$z_s = \frac{\alpha_3}{\sqrt{\text{Var}(\alpha_3)}}$$

where $Var(\alpha_3) = \frac{6(n-2)}{(n+1)(n+3)}$.

The test statistic of the kurtosis (K) test is computed by

$$z_k = \frac{\alpha_4 - E(\alpha_4)}{\sqrt{Var(\alpha_4)}}$$

where $E(\alpha_4) = \frac{3(n-1)}{n+1}$ and $Var(\alpha_4) = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}$.

The JB test is defined as the sum of squares of standardized skewness and kurtosis coefficients. The test is computed as

$$JB = n \left(\frac{\alpha_3^2}{6} + \frac{(\alpha_4^2 - 3)^2}{24} \right)$$

where n is the sample size. The JB test statistic equals zero when the distribution has zero skewness and kurtosis is 3. The large value of JB test statistic are lead to the rejection of the normality.

DA test just like the JB test, D'Agostino and Pearson also combined the skewness and kurtosis tests by simply taking the sum of the squares of the skewness and kurtosis represented as

$$DA = Z^2(\alpha_3) + Z^2(\alpha_4)$$

where $Z(\alpha_3)$ and $Z(\alpha_4)$ are the standardized normal equivalent deviates. $Z(\alpha_3)$ is defined by

$$Z(\alpha_3) = \delta \ln \left(\frac{Y}{a} + \sqrt{\frac{Y^2}{a^2} + 1} \right)$$

where δ , Y and a constants are computed by following equations $Y = \alpha_3 \left(\frac{(n+1)(n+3)}{6(n-2)} \right)^{1/2}$, $\delta = 1/\sqrt{\ln W}$

and $a = \left(\frac{2}{W^2 - 1} \right)^{1/2}$. In the equations, β_1 and W are computed as respectively $\beta_1 = \frac{3(n^2 + 27n - 70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)}$ and $W^2 = -1 + (2\beta_1 - 1)^{1/2}$. In the DA test, $Z(\alpha_4)$ is defined as

$$Z(\alpha_4) = \frac{1 - 2/9A - \left\{ (1 - 2/A) / (1 + z_k \sqrt{2/(A-4)}) \right\}^{1/3}}{\sqrt{2/9A}}$$

where $A = 6 + \frac{8}{\beta_2} \left(\frac{2}{\beta_2} + \sqrt{1 + \frac{4}{\beta_2}} \right)$ and $\beta_2 = \frac{6(n^2 - 5n + 2)}{(n+7)(n+9)} \sqrt{\frac{6(n+3)(n+5)}{n(n-2)(n-3)}}$. The test statistic DA is approximately distributed as chi-square with two degrees of freedom when the population is normally distributed.

Another normality test AJB test statistic is defined by following equation.

$$AJB = \frac{\alpha_3^2}{Var(\alpha_3)} + \frac{(\alpha_4 - E(\alpha_4))^2}{Var(\alpha_4)}$$

The Pearson chi-square (PC), Hegazy and Green (HG1, HG2), Geary (G) and Spiegelhalter's (Sp) statistics are chi-square and similar tests as. These are normality tests based on differences between observed and expected values. The large value of test statistics is lead to the rejection of normality.

For the PC normality test, the classified frequency series is created from the data and test statistic is defined as

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

where k , is the number of classes, O_i is the observed frequency for i th bin and E_i is the expected frequency of i th bin. The expected frequency is computed as $E_i = n[F(X_u) - F(X_l)]$ where X_u and X_l being the upper and lower limits for class i , respectively. $F(\cdot)$ is denoted by cumulative distribution function of normal distribution.

The proposed normality tests by Hegazy and Green [10] are based on difference between standardized values and expectation values of i th order statistics according to standard normal distribution. The HG1 and HG2 test statistics are calculated as follows

$$HG_1 = \frac{1}{n} \sum_{i=1}^n |z_i - \eta_i|$$

$$HG_2 = \frac{1}{n} \sum_{i=1}^n (z_i - \eta_i)^2$$

where $z_i = \frac{X_{(i)} - \bar{X}}{s}$ and $\eta_i = \Phi^{-1}\left(\frac{i}{n+1}\right)$.

The G test statistic for normality is defined as

$$G = \frac{1}{ns} \sum_{i=1}^n |X_i - \bar{X}|$$

where $s = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ is sample standard deviation.

The Sp statistic is a combination of Geary's and David-Hartley-Pearson's statistics

$$Sp = \left\{ \frac{1}{2n} (n!)^{\frac{1}{n-1}} \frac{R}{s} + \left(\frac{G}{\sqrt{\frac{n-1}{n}}} \right)^{-(n-1)} \right\}^{\frac{1}{n-1}}$$

where $R = X_{max} - X_{min}$ is range of sample.

3. R PROGRAMMING

R is a programming language developed for statistical computing and graphics. The R is a GNU project similar to the S, and although it is considered a different application of S, most code written in S language runs correctly in R.

R provides a wide range of statistical (classical tests, classification, linear and nonlinear modeling, time series analysis, clustering etc.) and graphical techniques. R is an extensible language. The S is often the preferred tool for research in statistical methodology and R provides an open source way to participate in this activity.

Total 18 Normality tests are built in four different packages: stats, nortest, normtest and mo-ments in R. KS and SW tests are in stats package. AD, CM, PC, SF and LF tests are in nortest package. AJB, B, G, HG1, HG2, JB, K, S, Sp and WB tests are in normtest package. DA test is in moments package.

The coding is quite easy in R. For example, AD normality test for random sample generated from lognormal distribution can be obtained by below codes in R.

```
library(nortest)
x=rlnorm(n,a,b)
ad.test(x)$p.value
```

Here n denotes sample size, a and b are parameters of lognormal distribution. If nortest package is not included in R it can be added by use of a command `install.packages("nortest")`.

4. APPLICATION

Simulation study designed to compare empirical type 1 error rate (EER) and power of 18 normality tests. Firstly, random samples are generated from normal distribution according to different sample sizes and parameter values in order to investigate the consistency of EERs. Then the power comparisons of normality tests are done respect to random samples generated from symmetric and asymmetric non-normal distributions given in Figure 1. The five symmetric distributions were $t(1)$, $t(3)$, Logistic(0,1), Cauchy(0,2), Cauchy(0,0.5), Laplace(0,1). The three asymmetric distributions were Gumbel (1,0.5), Gamma(2,1), Lognormal(1,0.5). The number of repetitions is taken 100,000 for each of these distributions.

The proportion of the smaller p values than α assumed as EER. The smallest value of the absolute difference between actual type 1 error(α) and EER indicates the best performance. The EERs of normality tests are given in Table 1.

According to Table 1, the KS test showed the worst performance in terms of the consistency of EER. Error plots of the EERs for normality tests except KS are given in Figure 2. It can be seen that from Figure 2, SF test has bigger EER than $\alpha=0.05$ for all cases.

The power of normality tests for random samples generated from non-normal symmetric and asymmetric distributions are computed according to different sample sizes and results are given in Table 2 and 3 respectively.

In the conducted simulation study with using non-normal symmetric distributions, HG2 test has the best power among all normality tests. As expected, the differences between the power of the normality tests have been decreased when sample size is increased. The Geary test showed the worst power performance among all normality tests. Even though KS test has widespread using in the statistical analysis showed poor power performance especially for small sample size.

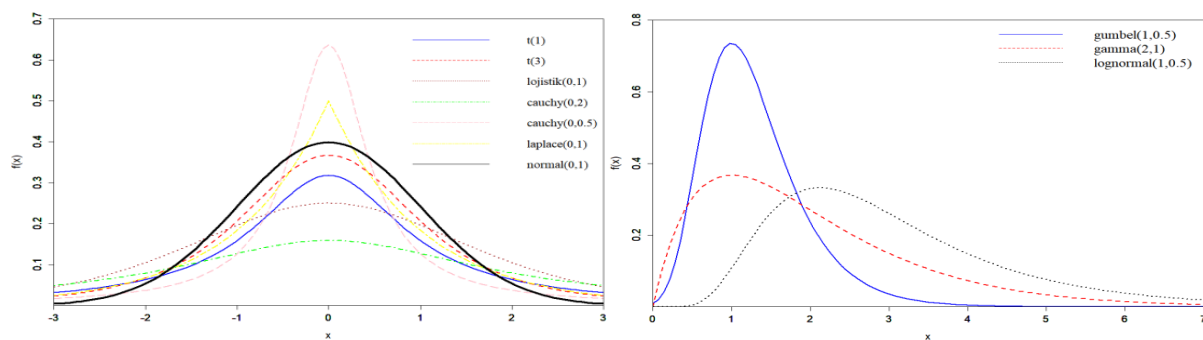


Figure 1. Plots of probability density functions of (a)symmetric and (b)asymmetric distributions

In the conducted simulation study with using non-normal asymmetric distributions, SW test is found the best normality test with regard to power of test. Also, DA test is showed better power performance than other normality tests.

Table 1. The empirical type-I error rates of normality tests according to different samples size and parameter ($\alpha = 0.05$)

	n	KS	SW	AD	CM	PC	SF	LF	AJB	B	G	HG1	HG2	JB	K	S	Sp	W	DA
N(0,0.01)	10	0.00002	0.04950	0.04971	0.04882	0.06643	0.05362	0.04955	0.04965	0.04989	0.05096	*0.05005	0.04913	0.04975	0.04934	0.04960	0.04916	0.04969	0.05220
	20	0.00005	0.04992	0.04959	0.05024	0.04957	0.05208	0.04795	*0.05001	0.05015	0.04996	0.04979	0.04990	0.04991	0.04970	0.04961	0.05026	0.04920	0.04985
	30	0.00013	0.05199	0.05102	0.05102	0.05139	0.05321	0.04924	0.05063	0.05119	0.05036	0.05099	0.05096	0.05081	0.05066	0.05120	*0.05010	0.05056	0.05096
	50	0.00025	0.04968	0.04894	0.04960	0.05230	0.05205	0.04797	0.04963	0.04962	0.04962	0.04923	*0.05015	0.04975	0.04947	0.04946	0.04962	0.04946	0.04907
	100	0.00011	0.04946	0.04928	0.05037	*0.05010	0.05151	0.05102	0.04897	0.04984	0.05102	0.05018	0.04938	0.04880	0.04940	0.04963	0.05066	0.04933	0.04900
N(0,0.25)	10	0.00001	0.04922	0.05006	0.04919	0.06606	0.05402	*0.05000	0.05064	0.05047	0.05056	0.04963	0.04965	0.05068	0.05037	0.05068	0.05042	0.04974	0.05302
	20	0.00004	0.05016	0.04959	0.05016	*0.04997	0.05284	0.04910	0.05012	0.04910	0.04896	0.04968	0.05033	0.05054	0.05019	0.05037	0.04957	0.05006	0.05073
	30	0.00013	0.05046	0.05019	0.05096	0.05086	0.05252	0.04902	0.04981	0.04993	0.05039	0.04985	0.05004	*0.05000	0.04979	0.05052	0.05004	0.05026	0.05064
	50	0.00019	0.05024	0.04955	*0.05003	0.05167	0.05234	0.04875	0.05075	0.04963	0.04976	0.05032	0.05025	0.05012	0.05059	0.04965	0.05099	0.05046	0.04920
	100	0.00019	*0.04996	0.04973	0.05098	0.05199	0.05234	0.05068	0.05075	0.05069	*0.04996	0.05053	0.05076	0.05080	0.05006	0.05079	0.04953	0.05092	0.05028
N(0,0.1)	10	0.00002	0.04927	0.04939	0.04833	0.06599	0.05364	0.04927	0.04935	*0.04983	0.05110	0.04969	0.04911	0.04954	0.04935	0.04892	0.05064	0.04920	0.05164
	20	0.00011	0.04993	0.04970	*0.04999	0.04826	0.05283	0.04739	0.05055	*0.04999	0.05123	0.04955	0.04981	0.05012	0.05147	0.04947	0.05068	0.04998	0.04985
	30	0.00013	0.05119	0.05071	0.05130	0.05051	0.05320	0.04830	0.05073	*0.05017	0.05043	0.05073	0.05083	0.05081	0.05142	0.05045	0.05143	0.05074	0.05027
	50	0.00010	0.05034	0.04970	0.05086	0.05208	0.05228	0.04899	0.04960	0.05011	0.04931	*0.05002	0.04957	0.04975	0.04965	0.04927	0.04973	0.05032	0.04901
	100	0.00011	0.04835	0.04948	0.05036	0.05038	0.05150	0.05023	0.04992	0.04965	0.04835	0.04956	0.04916	0.04990	0.04908	*0.05005	0.04993	0.04945	0.04978
N(0,1)	10	0.00002	0.05072	0.05111	0.04993	0.06683	0.05519	*0.04999	0.05033	0.05100	0.05052	0.05103	0.05074	0.05091	0.04936	0.05123	0.05055	0.05065	0.05366
	20	0.00008	0.05155	0.05053	0.05099	0.05026	0.05397	0.04852	0.05128	0.05106	*0.04996	0.05129	0.05131	0.05130	0.05084	0.05086	0.05076	0.05136	0.05107
	30	0.00015	0.05085	0.05040	0.05057	0.05138	0.05327	0.04860	0.05031	*0.05000	0.04977	0.04996	0.05070	0.05049	0.05049	0.05005	0.04934	0.05058	0.04971
	50	0.00015	0.04888	0.04882	0.04983	0.05202	0.05150	0.04866	0.05037	0.04914	0.04926	0.04889	0.04962	*0.04995	0.05041	0.04950	0.04938	0.04919	0.04900
	100	0.00017	0.04878	0.05055	0.05145	0.05335	0.05083	0.05133	0.04909	0.05120	0.04937	0.05051	0.04907	0.04864	0.04906	0.04929	*0.05022	0.04920	0.04904
N(0,10)	10	0.00002	*0.05003	0.05034	0.04930	0.06581	0.05462	0.04946	0.05028	*0.05003	0.05134	0.04976	0.05006	0.05040	0.05111	0.05032	0.05029	0.05041	0.05267
	20	0.00007	0.05027	0.05061	0.05132	0.04994	0.05260	0.04908	0.04919	0.05011	0.04947	*0.05004	0.04947	0.04923	0.04943	0.04959	0.05018	0.04968	0.05013
	30	0.00011	0.05089	*0.05001	0.05019	0.04975	0.05368	0.04865	0.05154	0.04981	0.04961	0.05030	0.05170	0.05136	0.05102	0.05122	0.05021	0.05133	0.05075
	50	0.00007	0.04973	0.04903	0.04942	0.05165	0.05249	0.04905	0.05019	0.04892	0.04899	0.04930	*0.05013	0.05053	0.04987	0.05085	0.05075	0.05027	0.05069
	100	0.00015	0.04982	0.04943	0.04996	0.05126	0.05173	0.05048	0.05004	0.04987	0.04886	0.04995	0.05012	0.05020	0.04995	0.04992	0.04873	*0.04999	0.04961
N(0,100)	10	0.00002	0.04916	0.04948	0.04913	0.06608	0.05457	0.04949	0.04997	0.04951	0.05009	0.04944	0.05022	0.04975	0.05013	0.05004	0.04988	*0.05002	0.05234
	20	0.00016	*0.04991	0.04939	0.05010	0.04903	0.05255	0.04805	0.04943	0.04974	0.04971	0.04987	0.04986	0.04948	0.04936	0.04957	0.04874	0.04970	0.05020
	30	0.00010	*0.05001	0.04961	0.05077	0.05117	0.05184	0.04762	0.04978	0.04998	0.05002	0.05036	0.04922	0.04938	0.05046	0.04890	0.05011	0.04960	0.04856
	50	0.00013	0.05098	0.05095	0.05152	0.05171	0.05293	0.04893	0.05098	0.05082	*0.05010	0.05114	0.05059	0.05079	0.05074	0.05031	0.04980	0.05107	0.04958
	100	0.00014	0.04954	0.04946	0.05071	0.05202	0.05183	0.05202	0.04963	0.05059	0.05132	0.05030	0.04968	0.04952	*0.04997	0.05018	0.04947	0.05005	0.05008

* Corresponding author. Email address: akarakoca@konya.edu.tr
<http://dergipark.gov.tr/csj> ©2016 Faculty of Science, Cumhuriyet University

Table 2. The power of the normality tests according to different sample sizes for the non-normal symmetric distributions.

	n	KS	SW	AD	CM	PC	SF	LF	AJB	B	G	HG1	HG2	JB	K	S	Sp	W	DA
Cauchy(0,0.5)	10	0.1690	0.5919	0.6134	0.6134	0.5356	0.6333	0.5770	0.5963	0.5924	0.0060	0.5975	0.6350	0.5876	0.5161	0.5744	0.5056	0.6256	0.5791
	20	0.5091	0.8642	0.8786	0.8766	0.7750	0.8904	0.8416	0.8692	0.8663	0.0005	0.8757	0.8958	0.8593	0.8563	0.7706	0.8603	0.8879	0.7716
	30	0.7294	0.9591	0.9654	0.9639	0.9091	0.9696	0.9446	0.9594	0.9604	0.0001	0.9652	0.9721	0.9546	0.9591	0.8462	0.9674	0.9688	0.8464
	50	0.9299	0.9967	0.9974	0.9973	0.9844	0.9979	0.9936	0.9963	0.9970	0.0000	0.9975	0.9981	0.9957	0.9965	0.9072	0.9986	0.9977	0.9072
	100	0.9988	1.0000	1.0000	1.0000	0.9999	1.0000	1.0000	1.0000	1.0000	0.0000	1.0000	1.0000	1.0000	1.0000	0.9537	1.0000	1.0000	0.9538
Cauchy(0,2)	10	0.1682	0.5919	0.6127	0.6127	0.5356	0.6350	0.5784	0.5952	0.5911	0.0057	0.5972	0.6367	0.5870	0.5143	0.5743	0.5056	0.6271	0.5797
	20	0.5131	0.8670	0.8806	0.8787	0.7790	0.8920	0.8422	0.8712	0.8682	0.0004	0.8776	0.8971	0.8619	0.8593	0.7719	0.8631	0.8894	0.7726
	30	0.7300	0.9585	0.9650	0.9641	0.9085	0.9691	0.9440	0.9581	0.9608	0.0000	0.9652	0.9719	0.9539	0.9576	0.8469	0.9666	0.9684	0.8470
	50	0.9289	0.9963	0.9972	0.9969	0.9842	0.9976	0.9935	0.9960	0.9967	0.0000	0.9974	0.9981	0.9954	0.9965	0.9070	0.9986	0.9975	0.9071
	100	0.9986	1.0000	1.0000	1.0000	0.9998	1.0000	1.0000	1.0000	1.0000	0.0000	1.0000	1.0000	1.0000	1.0000	0.9532	1.0000	1.0000	0.9530
Laplace (0,1)	10	0.0007	0.1517	0.1583	0.1550	0.1312	0.1852	0.1411	0.1830	0.1406	0.0177	0.1461	0.1844	0.1772	0.1212	0.1721	0.1311	0.1767	0.1773
	20	0.0050	0.2598	0.2725	0.2657	0.1439	0.3186	0.2156	0.3118	0.2411	0.0060	0.2657	0.3289	0.2992	0.2885	0.2540	0.2795	0.3113	0.2556
	30	0.0113	0.3555	0.3715	0.3640	0.1918	0.4272	0.2904	0.4130	0.3384	0.0021	0.3715	0.4454	0.3976	0.4078	0.3004	0.4375	0.4198	0.3006
	50	0.0260	0.5193	0.5439	0.5367	0.2732	0.5957	0.4277	0.5700	0.5156	0.0003	0.5566	0.6192	0.5530	0.5830	0.3538	0.6928	0.5896	0.3534
	100	0.0845	0.7962	0.8264	0.8234	0.4722	0.8444	0.7066	0.8089	0.8155	0.0000	0.8401	0.8612	0.7975	0.8337	0.4101	0.9381	0.8405	0.4102
Logistic(0,1)	10	0.0001	0.0808	0.0801	0.0758	0.0805	0.0966	0.0722	0.0986	0.0729	0.0346	0.0755	0.0933	0.0962	0.0701	0.0950	0.0692	0.0908	0.0988
	20	0.0006	0.1159	0.1052	0.0981	0.0649	0.1426	0.0821	0.1525	0.0900	0.0240	0.1026	0.1456	0.1484	0.1365	0.1328	0.1057	0.1373	0.1337
	30	0.0008	0.1473	0.1256	0.1145	0.0727	0.1833	0.0928	0.1966	0.1073	0.0182	0.1278	0.1902	0.1912	0.1857	0.1607	0.1475	0.1784	0.1602
	50	0.0013	0.1979	0.1604	0.1437	0.0779	0.2473	0.1120	0.2661	0.1363	0.0108	0.1705	0.2609	0.2587	0.2641	0.1902	0.2362	0.2417	0.1896
	100	0.0017	0.3022	0.2361	0.2086	0.0876	0.3703	0.1547	0.4002	0.2034	0.0034	0.2631	0.3964	0.3904	0.4099	0.2252	0.4069	0.3651	0.2247
t(1)	10	0.1679	0.5883	0.6085	0.6089	0.5348	0.6302	0.5756	0.5918	0.5881	0.0059	0.5939	0.6310	0.5832	0.5108	0.5706	0.5023	0.6219	0.5755
	20	0.5104	0.8659	0.8803	0.8786	0.7774	0.8927	0.8427	0.8719	0.8686	0.0004	0.8777	0.8976	0.8616	0.8597	0.7724	0.8629	0.8902	0.7728
	30	0.7296	0.9590	0.9649	0.9640	0.9072	0.9693	0.9439	0.9589	0.9604	0.0000	0.9651	0.9715	0.9549	0.9579	0.8474	0.9681	0.9685	0.8473
	50	0.9309	0.9966	0.9974	0.9973	0.9850	0.9977	0.9936	0.9958	0.9972	0.0000	0.9975	0.9980	0.9954	0.9962	0.9077	0.9984	0.9976	0.9077
	100	0.9987	1.0000	1.0000	1.0000	0.9999	1.0000	1.0000	1.0000	1.0000	0.0000	1.0000	1.0000	1.0000	1.0000	0.9516	1.0000	1.0000	0.9517

Table 2 cont. The power of the normality tests according to different sample sizes for the non-normal symmetric distributions

	n	KS	SW	AD	CM	PC	SF	LF	AJB	B	G	HG1	HG2	JB	K	S	Sp	W	DA
t(3)	10	0.0061	0.1868	0.1869	0.1798	0.1506	0.2176	0.1607	0.2210	0.1679	0.0227	0.1745	0.2157	0.2180	0.1626	0.2130	0.1542	0.2094	0.2176
	20	0.0607	0.3599	0.3480	0.3292	0.2048	0.4090	0.2800	0.4195	0.3128	0.0397	0.3391	0.4177	0.4092	0.3961	0.3663	0.3475	0.4028	0.3672
	30	0.0565	0.4584	0.4332	0.4077	0.2354	0.5164	0.3381	0.5259	0.3909	0.0042	0.4337	0.5292	0.5148	0.5168	0.4243	0.4774	0.5103	0.4243
	50	0.1060	0.6384	0.6058	0.5734	0.3311	0.6968	0.4803	0.7050	0.5630	0.0012	0.6173	0.7133	0.6949	0.7091	0.5213	0.7076	0.6921	0.5210
	100	0.2398	0.8779	0.8521	0.8268	0.5322	0.9076	0.7371	0.9107	0.8241	0.0000	0.8689	0.9177	0.9060	0.9191	0.6406	0.9307	0.9057	0.6401

Table 3. The power of the normality tests according to different sample sizes for the non-normal asymmetric distributions.

	n	KS	SW	AD	CM	PC	SF	LF	AJB	B	G	HG1	HG2	JB	K	S	Sp	W	DA
Gumbel(2,1)	10	0.0007	0.2388	0.2209	0.2042	0.1983	0.2403	0.1689	0.1847	0.2121	0.0594	0.2215	0.2222	0.2058	0.1426	0.2191	0.0996	0.2300	0.2250
	20	0.0081	0.5302	0.4655	0.4194	0.2783	0.5008	0.3188	0.3634	0.4335	0.0532	0.4813	0.4614	0.4060	0.2601	0.4689	0.1537	0.4916	0.4714
	30	0.0236	0.7487	0.6606	0.5982	0.4109	0.7079	0.4609	0.5241	0.6153	0.0475	0.6858	0.6588	0.5773	0.3492	0.6689	0.1984	0.7001	0.6693
	50	0.0808	0.9487	0.8906	0.8357	0.6519	0.9258	0.6924	0.7753	0.8501	0.0360	0.9091	0.8918	0.8207	0.4844	0.8925	0.2669	0.9223	0.8926
	100	0.3345	0.9997	0.9973	0.9914	0.9486	0.9993	0.9545	0.9912	0.9932	0.0187	0.9983	0.9981	0.9949	0.7049	0.9970	0.3827	0.9992	0.9970
Gumbel(1,0.5)	10	0.0005	0.1505	0.1430	0.1328	0.1258	0.1603	0.1163	0.1416	0.1344	0.0475	0.1376	0.1504	0.1501	0.1073	0.1560	0.0804	0.1524	0.1608
	20	0.0038	0.3115	0.2726	0.2478	0.1409	0.3121	0.1982	0.2657	0.2491	0.0406	0.2720	0.2941	0.2880	0.2009	0.3216	0.1215	0.3053	0.3231
	30	0.0092	0.4607	0.3953	0.3581	0.1885	0.4521	0.2823	0.3847	0.3617	0.0344	0.4033	0.4261	0.4132	0.2756	0.4728	0.1601	0.4454	0.4734
	50	0.0258	0.6903	0.6018	0.5486	0.2763	0.6737	0.4370	0.5834	0.5552	0.0254	0.6234	0.6379	0.6169	0.3882	0.7030	0.2291	0.6675	0.7023
	100	0.1045	0.9412	0.8873	0.8441	0.4997	0.9316	0.7301	0.8882	0.8539	0.0131	0.9073	0.9126	0.9055	0.5865	0.9469	0.3640	0.9297	0.9473
LogN(1,0.5)	10	0.0014	0.2457	0.2300	0.2150	0.1965	0.2538	0.1791	0.2101	0.2197	0.0488	0.2271	0.2378	0.2295	0.1572	0.2411	0.1087	0.2425	0.2468
	20	0.0144	0.5220	0.4665	0.4289	0.2720	0.5092	0.3396	0.4093	0.4350	0.0382	0.4741	0.4799	0.4468	0.3076	0.4999	0.1850	0.5012	0.5015
	30	0.0370	0.7223	0.6505	0.6001	0.3834	0.6989	0.4798	0.5730	0.6112	0.0302	0.6684	0.6623	0.6156	0.4166	0.6919	0.2507	0.6918	0.6919
	50	0.1147	0.9243	0.8723	0.8299	0.5876	0.9073	0.7093	0.8104	0.8399	0.0187	0.8889	0.8811	0.8443	0.5820	0.9022	0.3711	0.9043	0.9020
	100	0.3955	0.9987	0.9944	0.9876	0.8885	0.9979	0.9509	0.9900	0.9893	0.0069	0.9961	0.9960	0.9936	0.8116	0.9973	0.5755	0.9978	0.9973

Table 4. The average powers of the normality tests according to different sample sizes in the non-normal symmetric and asymmetric distributions

Case	n	KS	SW	AD	CM	PC	SF	LF	AJB	B	G	HG1	HG2	JB	K	S	Sp	W	DA
Symmetric	10	0.0853	0.3652	0.3766	0.3743	0.3280	0.3996	0.3508	0.3810	0.3588	0.0154	0.3641	0.3993	0.3749	0.3159	0.3666	0.3113	0.3919	0.3713
	20	0.2665	0.5555	0.5609	0.5545	0.4575	0.5909	0.5174	0.5827	0.5412	0.0118	0.5564	0.5971	0.5732	0.5661	0.5113	0.5532	0.5865	0.5123
	30	0.3763	0.6396	0.6376	0.6297	0.5374	0.6725	0.5923	0.6686	0.6197	0.0041	0.6381	0.6800	0.6611	0.6641	0.5710	0.6607	0.6690	0.5710
	50	0.4872	0.7242	0.7170	0.7075	0.6060	0.7555	0.6668	0.7549	0.7010	0.0020	0.7228	0.7646	0.7488	0.7576	0.6312	0.7720	0.7527	0.6310
	100	0.5537	0.8294	0.8191	0.8098	0.6819	0.8537	0.7664	0.8533	0.8072	0.0006	0.8287	0.8625	0.8490	0.8604	0.6891	0.8793	0.8519	0.6889
Asymmetric	10	0.0009	0.2117	0.1980	0.1840	0.1735	0.2181	0.1548	0.1788	0.1887	0.0519	0.1954	0.2035	0.1951	0.1357	0.2054	0.0962	0.2083	0.2109
	20	0.0088	0.4546	0.4015	0.3654	0.2304	0.4407	0.2855	0.3461	0.3725	0.0440	0.4091	0.4118	0.3803	0.2562	0.4301	0.1534	0.4327	0.4320
	30	0.0233	0.6439	0.5688	0.5188	0.3276	0.6196	0.4077	0.4939	0.5294	0.0374	0.5858	0.5824	0.5354	0.3471	0.6112	0.2031	0.6124	0.6115
	50	0.0737	0.8544	0.7882	0.7381	0.5053	0.8356	0.6129	0.7230	0.7484	0.0267	0.8071	0.8036	0.7606	0.4849	0.8326	0.2890	0.8314	0.8323
	100	0.2782	0.9799	0.9597	0.9410	0.7789	0.9763	0.8785	0.9564	0.9455	0.0129	0.9672	0.9689	0.9647	0.7010	0.9804	0.4407	0.9756	0.9805
Total	10	0.0572	0.3140	0.3171	0.3108	0.2765	0.3391	0.2855	0.3136	0.3021	0.0276	0.3079	0.3341	0.3149	0.2558	0.3128	0.2396	0.3307	0.3179
	20	0.1806	0.5218	0.5078	0.4914	0.3818	0.5408	0.4401	0.5038	0.4849	0.0226	0.5073	0.5353	0.5089	0.4628	0.4843	0.4199	0.5352	0.4855
	30	0.2586	0.6410	0.6147	0.5927	0.4675	0.6549	0.5308	0.6104	0.5896	0.0152	0.6207	0.6475	0.6192	0.5585	0.5844	0.5082	0.6502	0.5845
	50	0.3493	0.7676	0.7407	0.7177	0.5724	0.7822	0.6488	0.7443	0.7168	0.0103	0.7509	0.7776	0.7528	0.6667	0.6983	0.6110	0.7789	0.6981
	100	0.4618	0.8795	0.8660	0.8535	0.7143	0.8946	0.8038	0.8877	0.8533	0.0047	0.8749	0.8980	0.8875	0.8073	0.7862	0.7331	0.8931	0.7861

The power averages for symmetric and asymmetric distributions according to different sample sizes are given in Table 4. Also, the error plots of power averages for symmetric and asymmetric distributions are given in Figure 3 as generally.

Considering all situations in the simulation study, SF test is the best normality test in terms of power for sample sizes 10, 20, 30 and 50. The HG2 test is the best normality test for sample size 100. In terms of power for symmetric distributions, SF(for sample size 10), HG2(for sample sizes 20 and 30) and Sp(for sample sizes 50 and 100) are found the most successful tests. In the non-normal asymmetric distributions, SF(for sample size 10), SW(for sample sizes 20, 30 and 50) and DA(for sample size 100) are found the most successful normality tests according to power of test.

The investigated normality tests except SE, S, DA and G showed better performance for the non-normal symmetric than asymmetric distributions.

HG2 test for non-normal symmetric distributions and SW test for non-normal asymmetric distributions are found the most successful normality test when power of the test without considering sample size is compared.

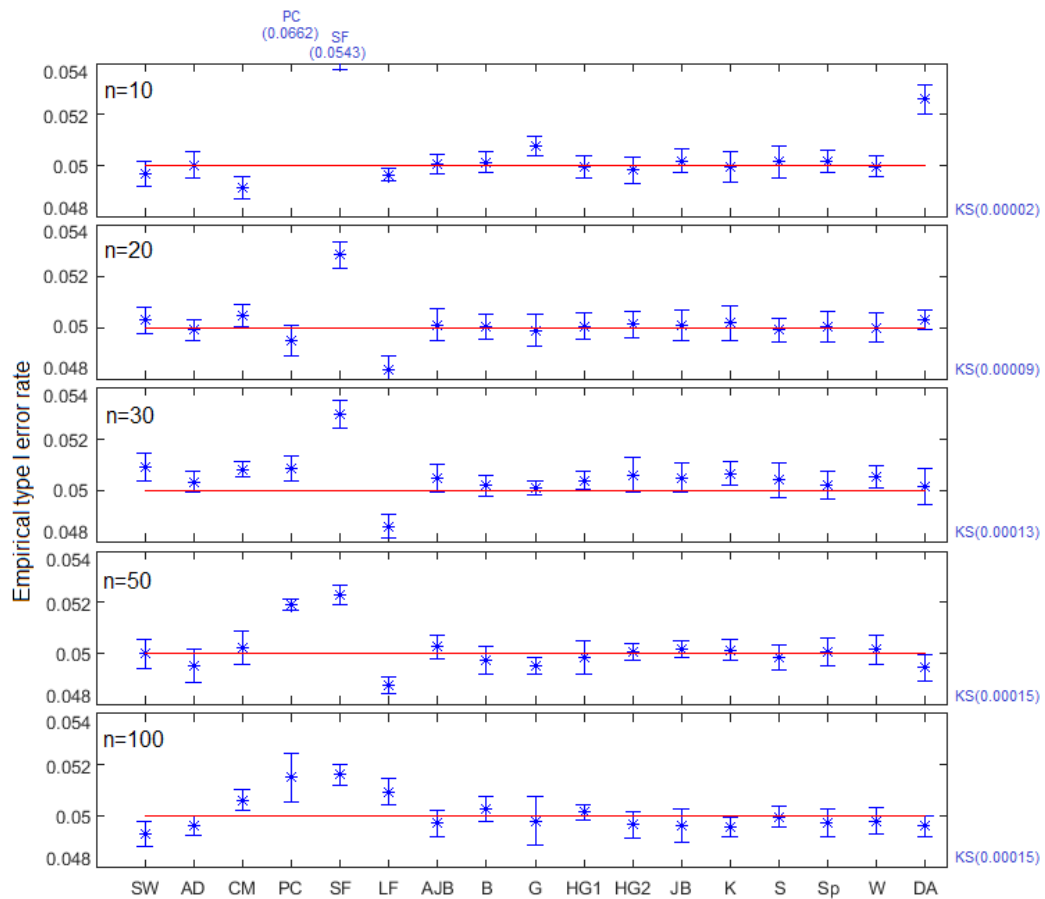


Figure 2. Error plots of empirical type 1 error rates for normality tests according to different sample size.

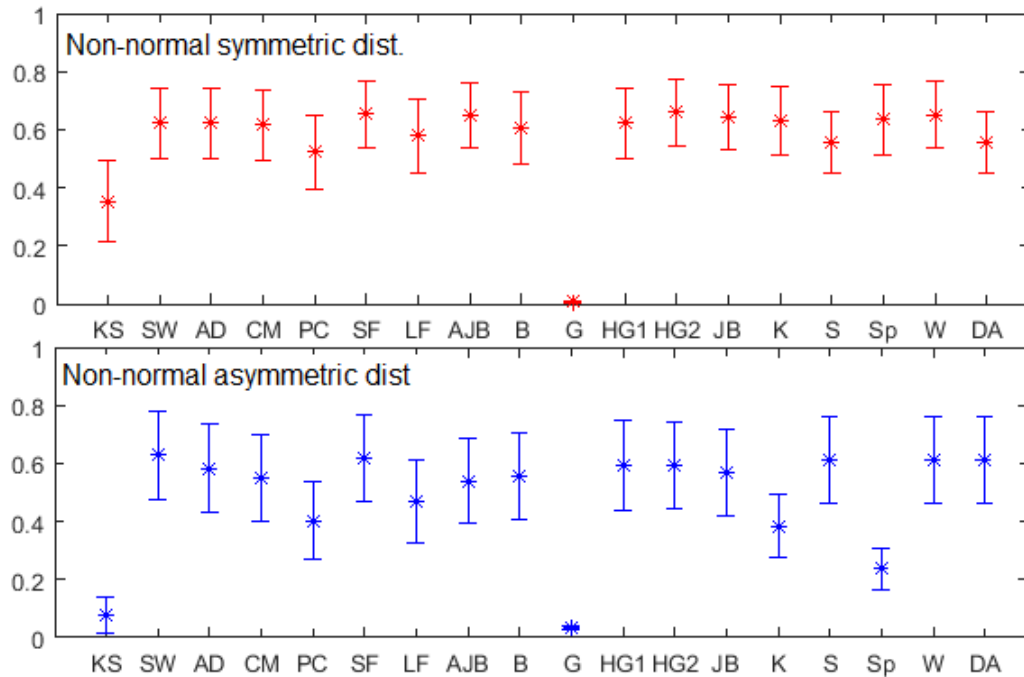


Figure 3. Error plots of test powers for normality tests as generally.

5. CONCLUSIONS

In this study, the normality tests in the free software R are compared according to consistency of type 1 error rate and test power. The comparison results can be summarized as follows:

- The widely used KS test showed worse performance than other normality tests in terms of both test power and consistency in the comparison of normality tests.
- The Frosini's B test showed better performance than other normality tests according to consistency of type 1 error rate when evaluated generally.
- The SF test should be used when there is no information about the shape of distribution.
- The SF and HG2 test should be used in terms of the test power for the small sample sizes in the symmetric distributions.
- The Sp test should be used in terms of the test power for the large sample sizes in the symmetric distributions.
- The SW test should be used in terms of the test power in the asymmetric distributions.

REFERENCES

- [1] Pearson, K., X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 50-302 (1900) 157–175.
- [2] Von Mises, R., *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik*. Bull. Amer. Math. Soc 38 (1932) 169-170.
- [3] Kolmogorov, A., *Sulla determinazione empirica di una legge di distribuzione*, *Inst. Ital. Attuari, Giorn.* 4 (1933) 83–91.
- [4] Geary, R. C., *The ratio of the mean deviation to the standard deviation as a test of normality*, *Biometrika* 27-3/4 (1935) 310–332.
- [5] Anderson, T. W. and Darling, D. A., *A test of goodness of fit*, *Journal of the American statistical association* 49-268 (1954) 765–769.
- [6] Kuiper, N. H., *Tests concerning random points on a circle*, *Nederl. Akad. Wetensch. Proc. Ser. A*, Vol. 63, (1960) 38–47.

- [7] Shapiro, S. S. and Wilk, M. B., An analysis of variance test for normality (complete samples), *Biometrika* 52-3/4 (1965) 591–611.
- [8] Lilliefors, H. W., On the kolmogorov-smirnov test for normality with mean and variance unknown, *Journal of the American statistical Association* 62-318 (1967) 399–402.
- [9] Shapiro, S. S. and Francia, R., An approximate analysis of variance test for normality, *Journal of the American Statistical Association* 67-337 (1972) 215–216.
- [10] Hegazy, Y. and Green, J., Some new goodness-of-fit tests using order statistics, *Applied Statistics* pp. (1975) 299–308.
- [11] Weisberg, S. and Bingham, C., An approximate analysis of variance test for non-normality suitable for machine calculation, *Technometrics* 17-1 (1975) 133–134.
- [12] Spiegelhalter, D., A test for normality against symmetric alternatives, *Biometrika* 64-2 (1977) 415–418.
- [13] Frosini, B. V., A survey of a class of goodness-of-fit statistics, *Universit`a degli studi. Facolt`a di scienze statistiche, demografiche ed attuariali* (1978).
- [14] Jarque, C. M. and Bera, A. K., A test for normality of observations and regression residuals, *International Statistical Review/Revue Internationale de Statistique* (1987) 163–172.
- [15] Urzua, C. M., On the correct use of omnibus tests for normality, *Economics Letters* 3-54 (1997) 301.
- [16] Esteban, M., Castellanos, M., Morales, D. and Vajda, I., Monte carlo comparison of four normality tests using different entropy estimates, *Communications in Statistics-Simulation and computation* 30-4 (2001) 761–785.
- [17] Yazici, B. and Yolacan, S., A comparison of various tests of normality, *Journal of Statistical Computation and Simulation* 77-2 (2007) 175–183.
- [18] Yap, B. W. and Sim, C. H., Comparisons of various types of normality tests, *Journal of Statistical Computation and Simulation* 81-12 (2011) 2141–2155.
- [19] Noughabi, H. A. and Arghami, N. R., Monte carlo comparison of seven normality tests, *Journal of Statistical Computation and Simulation* 81-8 (2011) 965–972.
- [20] Royston, J., An extension of shapiro and wilk's w test for normality to large samples, *Applied statistics* pp. (1982) 115–124.
- [21] Razali, N. M., & Wah, Y. B., Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of statistical modeling and analytics*, 2-1(2011) 21-33.