International Journal of Assessment Tools in Education
2022, Vol. 9, No. 2, 397–409
https://doi.org/10.21449/ijate.1101295
Published at https://ijate.net/ https://dergipark.org.tr/en/pub/ijate **Research Article**

Comparison of Normality Tests in Terms of Sample Sizes under Different Skewness and Kurtosis Coefficients

Suleyman Demir[1,*]

[1]Sakarya University, Faculty of Education, Department of Educational Sciences, Sakarya, Türkiye.

ARTICLE HISTORY

Received: Apr. 06, 2021
Revised: Feb. 14, 2022
Accepted: Apr. 05, 2022

Keywords:
Normal distribution,
Skewness,
Kurtosis,
Normality tests.

Abstract: This study aims to compare normality tests in different sample sizes in data with normal distribution under different kurtosis and skewness coefficients obtained simulatively. To this end, firstly, simulative data were produced using the MATLAB program for different skewness/kurtosis coefficients and different sample sizes. The normality analysis of each data type was conducted using the MATLAB program and ten different normality tests; namely, (Kolmogorov Smirnov (KS) Test, KS Stephens Modification, KS Marsaglia, KS Lilliefors Modification, Anderson-Darling Test, Cramer- Von Mises Test, Shapiro-Wilk Test, Shapiro-Francia Test, Jarque-Bera Test, and D'Agostino & Pearson Test). As a result of the analyses conducted according to ten different normality tests, it was found that normality tests were not affected by the sample size when the skewness and kurtosis coefficients were equal to or close to zero; however, in cases where the skewness and kurtosis coefficients moved away from zero, it was found that normality tests are affected by the sample size, and such tests tend to give significant results. Therefore, in large samples, it may be suggested that critical values for skewness and kurtosis coefficients' z-scores as proposed by Kim (2013) and Mayers (2013) or the histogram graphs be used.

1. INTRODUCTION

Parametric methods in data analysis (t-tests, ANOVA, etc.) are used in cases where the data obtained from the sample have a normal distribution. If the data do not have a normal distribution, non-parametric methods (Mann Whitney U, Kruskal Wallis, Wilcoxon, etc.) are used to analyze data. Parametric methods, based on a specific distribution such as normal distribution, can be used in conditions where the normality assumption is provided. Non-parametric methods are implemented independently of the distribution, converting data to ordinal data type (Field, 2013).

The definition of normal distribution was first introduced by Abraham de Moivre in 1667 and defined by a mathematical formula (Howell, 2013; Martin & Bridgmon, 2012). Also called Gaussian distribution as given in Formula 1.1. If it is accepted as ($\mu = 0$) and ($\sigma = 1$) in the formula in 1.1, the standard normal distribution function is expressed as (1.2).

*CONTACT: Suleyman DEMIR suleymand@sakarya.edu.tr Sakarya University, Faculty of Education, Department of Educational Sciences, Sakarya, Türkiye

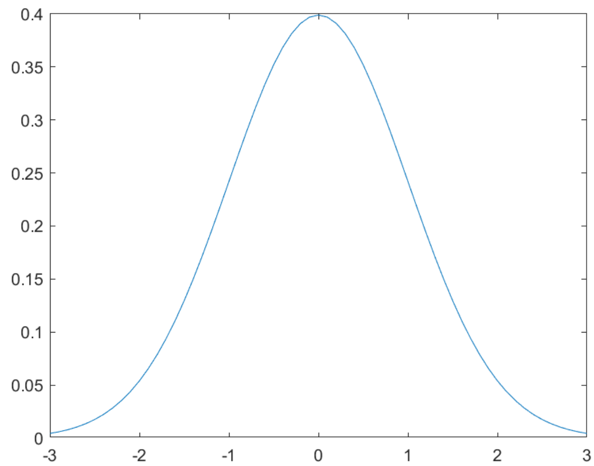
e-ISSN: 2148-7456 /© **IJATE** 2022

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{1.1}$$

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}x^2} \tag{1.2}$$

The standard normal distribution chart given in Figure 1 has standard deviation (σ) values on the X-axis, while probability values are on the Y-axis. For a sample with a standard normal distribution; 68.2% of the sample fall within the ±1 standard deviation range, 95.4% within ±2 standard deviations, 99.7% within ±3 standard deviations, and 0.3% outside of the ±3 standard deviation range (Field, 2013; Howell, 2013; Martin & Bridgmon, 2012; Tabachnick & Fidell, 2013).

Figure 1. *Standard normal distribution chart*



The normal distribution has two components, namely, skewness and kurtosis. Skewness is related to the status of the data's mode median and mean relative to each other. There is symmetric distribution when the mean is in the middle of the distribution; thus, there is no skewness. When the mean is not in the middle of the distribution, there is a non-symmetric distribution (skewed distribution) (Tabachnick & Fidell, 2007). The kurtosis is related to how far the data move away from the mean or how close they get to the mean. In other words, it is related to the standard deviation of data. When the standard deviation is small, there is a pointed distribution (leptokurtic, short-tailed); whereas, when the standard deviation is large, there is a flattened distribution (platykurtic, long-tailed) (Baykul & Güzeller, 2013; Field, 2013; Tabachnick & Fidell, 2013).

Although there are different methods for the calculation of the coefficient of skewness, Baykul and Güzeller (2013) state that using the mean as the central tendency measure of the coefficient of skewness and the standard deviation as the measure of the central distribution, thereby calculating it according to the third moment around the mean gives better results. Accordingly, the formula calculates the skewness coefficient in (1.3) according to the third moment around the mean. When the skewness coefficient is equal to zero, as in a normal distribution, there is symmetric distribution since the majority of the data are around the average. The skewness coefficient being negative indicates that most of the data are located on the right side of the mean and the tail on the left side is longer, while the skewness coefficient being positive shows that the majority of the data are located on the left side of the average and the tail on the right side is longer (Tabachnick & Fidell, 2013).

$$skew = \frac{(n-1)(n-2)}{n} \sum_{j=1}^{n} \frac{\left(X_j - \overline{X}\right)^3}{\sigma_x^3} \tag{1.3}$$

The formula calculates the kurtosis coefficient in (1.4) according to the fourth moment around the mean (formula in 1.5 is mostly preferred). A positive kurtosis coefficient indicates that the distribution is more pointed than the normal distribution. In contrast, a negative kurtosis coefficient indicates that the distribution is more flattened than the normal distribution. That the kurtosis coefficient equals zero indicates the distribution neither too flattened nor too pointed, as in the normal distribution (Field, 2013; Howell, 2013; Martin & Bridgmon, 2012; Tabachnick & Fidell, 2013).

$$kurt = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{j=1}^{n} \frac{\left(X_j - \overline{X}\right)^4}{\sigma_x^4} \tag{1.4}$$

$$kurt = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{j=1}^{n} \frac{\left(X_j - \overline{X}\right)^4}{\sigma_x^4} - 3 \tag{1.5}$$

The fact that both the skewness and kurtosis coefficients using the formulas in (1.3) and (1.5) are zero shows that the data have normal distribution; however, since the skewness and kurtosis values are mostly different from zero, acceptable ranges are determined for these values. These ranges have been suggested to be ±2 in some sources (Field, 2013; Gravetter & Wallnau, 2014; George & Mallery, 2010; Trochim & Donnelly, 2006); ±1.5 in some (Tabachnick & Fidell, 2013); and ±1 in other sources (Bulmer, 1979). In addition, Hair et al. (2010) and Bryne (2010) state that the normality assumption is not fulfilled when the skewness coefficient is outside the range of ±2 and the kurtosis coefficient is outside the range of ±7; while according to Kline (2011), these ranges are ±3 for the skewness coefficient and ±10 for the kurtosis coefficient.

In addition to skewness and kurtosis coefficients, standard z-scores of skewness and kurtosis coefficients, histogram graphs, and normality tests are also used to test the normality assumption. The standard z-score of the skewness coefficient is calculated as follows (1.6): the skewness coefficient is divided by the standard error value of the skewness coefficient (1.7). The standard z-score of the kurtosis coefficient is calculated as follows (1.6): the kurtosis coefficient is divided by the standard error value of the kurtosis coefficient (1.7). Field (2013) and Tabachnick and Fidell (2013) recognize that these values in the range of ±1.96 show a normal distribution in small samples. The standard z-scores fall outside the range of ±1.96 because the standard error values of the skewness and kurtosis coefficients decrease as the sample size increases. In such cases, the histogram graph should be interpreted instead of the standard z-scores (Field, 2013; Tabachnick & Fidell, 2013).

$$z_{skew} = \frac{skew - 0}{SE_{skew}} \tag{1.6}$$

$$SE_{skew} = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}} \tag{1.7}$$

$$z_{kurt} = \frac{kurt - 0}{SE_{kurt}} \tag{1.8}$$

$$SE_{kurt} = \sqrt{\frac{24n(n-1)^2}{(n-3)(n-2)(n+3)(n+5)}} \tag{1.9}$$

It was stated by Kim (2013) and Mayers (2013) that, for all sample sizes, the standard z-scores of the skewness and kurtosis coefficients which are within the range of ±1.96 are not sufficient for normal distribution. According to Kim (2013), the fact that when the sample number is n<50 and z-scores are in the range of ±1.96; and when the sample size is in the range of 50<n<300, and z-scores are in the range of ±3.29 means that the data have a normal distribution. When the sample size is n>300, the histogram graph should be interpreted. According to Mayers (2013), unlike Kim (2013), it is accepted that in cases where the sample size is in the range of 50<n<175, z-scores should be in the range of ±2.58.

Another method used to test the normality assumption is normality tests. The fact that normality tests are significant indicates that the data differ significantly from the normal distribution. In contrast, the fact that normality tests are not significant indicates that the data do not differ significantly from the normal distribution. In research studies, mostly Kolmogorov Smirnov (KS) Test (Smirnov, 1948) and Shapiro-Wilk Test (Shapiro & Wilk, 1965) are used. However, tests such as KS Stephens Modification, KS Marsaglia, KS Lilliefors Modification, Anderson-Darling Test, Cramer- Von Mises Test, Shapiro-Francia Test, Jarque-Bera Test, D'Agostino-Pearson Test (Detailed information is given in the Appendix) are also used (Lee et al., 2016; Marsaglia et al., 2003; Stephens, 1974).

Sample size affects the results of normality tests. When sample is small, normality tests tend to accept the null hypothesis. In large samples, even small deviations from the normal distribution cause the normality test to reject the null hypothesis (Öztuna et al., 2006). Some researchers (Lumley et al., 2002; Wilcox, 2010) state that in large samples, according to the central limit theorem, the data will approach normal distribution, therefore it can be assumed that the normal distribution assumption is achieved in large samples regardless of the normality determination methods However, some researchers (Micceri, 1989; Öztuna et al., 2006) state that this is not true: since the number of samples is large, the data will not always have a normal distribution. It seems more important to interpret the histogram graph with the skewness and kurtosis coefficients of the data.

Because of the confusion about the sample size, it is stated in the related literature that if the number of samples exceeds 200, the data should be considered to have a normal distribution due to the central limit theorem or that only the histogram graph should be interpreted. In the studies carried out, the power of the test was calculated for different sample sizes to determine the sensitivity of normality tests in data with normal and non-normal distribution (Douglas & Edith, 2002; Frain, 2007; Keskin, 2006; Nornadiah & Yap 2011; Nor-Aishah & Shamsul 2007; Öztuna et al., 2006; Rinnakorn & Kamon 2007; Stephens, 1974; Ukponmwan & Ajibade, 2017; Yap & Sim 2011). In cases where the sample size exceeds 200, it can be accepted that the data have a normal distribution or the histogram graph can be interpreted. However, the histogram graph will visually move away from the normal distribution as the kurtosis and skewness coefficients move away from zero. It is important to determine how much the histogram graph differs from the normal distribution according to the sample size, kurtosis coefficient, and skewness coefficient. However, when the relevant literature is reviewed, no study can be found to determine the sensitivity of normality tests under different sample sizes in the data that are considered to show the normal distribution in terms of different skewness and kurtosis coefficients. This study, therefore, aims to compare normality tests in terms of different sample sizes in data with normal distribution in terms of different kurtosis and skewness coefficients obtained simulatively.

2. METHOD

2.1. Obtaining Data

In order to realize the purpose of the study, firstly conditions were created depending on the kurtosis and skewness coefficients and the sample size. If the sample is less than 30, it is expressed as a small sample, and if the sample size is larger than 400, it is expressed as a large sample (Abbott, 2011, Demir et al., 2016; Orcan, 2020). In the present study, 11 different sample sizes were determined to cover small and large samples (10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 900). That the coefficient of kurtosis and skewness are zero means normal distribution. According to all the references in the related literature, the acceptable range for the skewness and kurtosis coefficients is ±1. In the current study, five different skewness and kurtosis values were determined, (-0.50, -0.25, 0, 0.25, 0.50) with a specific purpose to compare normality tests on data showing normal distribution (Bryne, 2010; Bulmer, 1979; Field, 2013; Gravetter & Wallnau, 2014; George & Mallery, 2010; Hair et al., 2010; Kline, 2011; Tabachnick & Fidell, 2013; Trochim & Donnelly, 2006). The sample sizes and skewness/kurtosis coefficients used in this study are shown in Table 1.

Table 1. *Sample sizes and skewness/kurtosis coefficients used for data generation*

	Sample Sizes	Skewness/Kurtosis Coeff.
Condition	10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 900	-0.50, -0.25, 0, 0.25, 0.50

The current study examined normality tests under 55 different conditions, including 11 different sample sizes and five different skewness/kurtosis coefficients.

2.2. Analysis of the Data

The data obtained according to five different skewness and kurtosis coefficients and 11 different sample sizes were analyzed using the MATLAB program according to 10 different normality tests (Kolmogorov Smirnov, KS Stephens Modification, KS Marsaglia, KS Lilliefors Modification, Anderson-Darling Test, Cramer- Von Mises Test, Shapiro-Wilk Test, Shapiro-Francia Test, Jarque-Bera Test, and D’Agostino & Pearson Test). MATLAB codes created by Öner and Kocakoç (2017) were used in the analysis phase.

3. FINDINGS

This section gives ten different normality test results of the data obtained simulatively under 55 different conditions. In Table 2, significance values obtained from normality tests are given for 11 different sample sizes under the condition that the kurtosis and skewness coefficients are -0.50. Table 2 shows that all normality tests do not have a normal distribution of data under conditions where the sample size is 200 or more for the significance level of α=0.05. Methods of Anderson-Darling Test and Cramer-Von Mises Test conclude that data do not have a normal distribution under conditions where the sample size is 40 or more for the significance level of α=0.05. Kolmogorov Smirnov, KS Marsaglia, Jarque-Bera Test, and D'Agostino & Pearson Test also conclude that data do not have a normal distribution under the sample size of 200 or more for the significance level of α=0.05. As a result, it can be said that the Anderson-Darling Test and Cramer-Von Mises Test methods are the most affected ones by the sample size when the coefficients of kurtosis and skewness are -0.50. At the same time, Kolmogorov Smirnov, KS Marsaglia, Jarque-Bera Test, and D’Agostino & Pearson Test are relatively less affected. In Table 3, the significance values obtained from the normality tests are given for 11 different sample sizes under conditions where the kurtosis and skewness coefficients are -0.25.

Table 2. *Normality test results for different samples in cases where kurtosis and skewness coefficients are (-0.50, -0.50)*

Sample Size	10	20	30	40	50	100	200	300	400	500	900
Normality Tests	*p*	*p*	*p*	*p*	*p*	*p*	*p*	*p*	*p*	*p*	*p*
Kolmogorov Smirnov	0.890	0.942	0.602	0.493	0.302	0.061	0.004	0.001	0.000	0.000	0.000
KS Stephens Modification	0.150	0.150	0.139	0.081	0.027	0.010	0.010	0.010	0.010	0.010	0.010
KS Marsaglia Method	0.832	0.911	0.556	0.455	0.276	0.056	0.004	0.001	0.000	0.000	0.000
KS Lilliefors Modification	0.200	0.200	0.140	0.079	0.019	0.000	0.000	0.000	0.000	0.000	0.000
Anderson-Darling Test	0.778	0.644	0.367	0.049	0.025	0.001	0.000	0.000	0.000	0.000	0.000
Cramer- Von Mises Test	0.751	0.735	0.327	0.045	0.020	0.000	0.000	0.000	0.000	0.000	0.000
Shapiro-Wilk Test	0.844	0.519	0.338	0.144	0.070	0.001	0.000	0.000	0.000	0.000	0.000
Shapiro-Francia Test	0.785	0.567	0.391	0.168	0.089	0.003	0.000	0.000	0.000	0.000	0.000
Jarque-Bera Test	0.771	0.594	0.458	0.353	0.272	0.074	0.005	0.000	0.000	0.000	0.000
D'Agostino & Pearson Test	0.652	0.545	0.427	0.329	0.253	0.065	0.004	0.000	0.000	0.000	0.000

Based on an analysis of Table 3, it can be concluded that all normality tests do not have a normal distribution of data under conditions where the sample size is 900 for the significance level of α=0.05. The KS Stephens Modification and KS Lilliefors Modification methods prove that data do not have a normal distribution under conditions where the sample size is 40 or more for the significance level of α=0.05. The Kolmogorov Smirnov and KS Marsaglia methods conclude that data do not have a normal distribution under conditions where the sample size is 900 for the significance level of α=0.05. In conclusion, it can be said that the KS Stephens Modification and KS Lilliefors Modification methods are affected most by the sample size when the coefficients of kurtosis and skewness are -0.25. At the same time, Kolmogorov Smirnov and KS Marsaglia are relatively less affected. In Table 4, the significance values obtained from the normality tests are given for 11 different sample sizes under conditions where the kurtosis and skewness coefficients are 0.00.

Table 3. *Normality test results for different samples in cases where kurtosis and skewness coefficients are (-0.25, -0.25)*

Sample Size	10	20	30	40	50	100	200	300	400	500	900
Normality Tests	*p*	*p*	*p*	*p*	*p*	*p*	*p*	*p*	*p*	*p*	*p*
Kolmogorov Smirnov	0.892	0.998	0.970	0.707	0.659	0.719	0.388	0.392	0.144	0.081	0.009
KS Stephens Modification	0.150	0.150	0.150	0.150	0.150	0.150	0.047	0.048	0.010	0.010	0.010
KS Marsaglia Method	0.835	0.993	0.953	0.666	0.622	0.693	0.372	0.379	0.139	0.078	0.008
KS Lilliefors Modification	0.200	0.200	0.200	0.200	0.200	0.200	0.046	0.049	0.003	0.001	0.000
Anderson-Darling Test	0.612	0.887	0.846	0.512	0.504	0.270	0.053	0.034	0.012	0.004	0.000
Cramer- Von Mises Test	0.475	0.878	0.774	0.362	0.358	0.217	0.072	0.046	0.012	0.002	0.000
Shapiro-Wilk Test	0.911	0.923	0.903	0.789	0.771	0.497	0.196	0.101	0.039	0.013	0.000
Shapiro-Francia Test	0.722	0.869	0.866	0.710	0.699	0.438	0.203	0.127	0.055	0.020	0.000
Jarque-Bera Test	0.937	0.878	0.823	0.771	0.722	0.522	0.272	0.142	0.074	0.039	0.003
D'Agostino & Pearson Test	0.763	0.822	0.802	0.767	0.728	0.538	0.280	0.144	0.074	0.037	0.003

Table 4 shows that all normality tests have a normal distribution of data for the significance level of α=0.05, regardless of the number of samples for the significance level of α=0.05. In conditions where the kurtosis and skewness coefficients are 0.00, it can be said that no normality test is affected by the sample size. In Table 5, the significance values obtained from the normality tests are given for 11 different sample sizes under conditions where the kurtosis and skewness coefficients are 0.25.

Table 4. *Normality test results for different samples in cases where kurtosis and skewness coefficients are (0.00, 0.00)*

Sample Size	10	20	30	40	50	100	200	300	400	500	900
Normality Tests	*p*	*p*	*p*	*p*	*p*	*p*	*p*	*p*	*p*	*p*	*p*
Kolmogorov Smirnov	0.846	0.968	0.953	0.838	0.858	0.922	0.770	0.733	0.716	0.777	0.527
KS Stephens Modification	0.150	0.150	0.150	0.150	0.150	0.150	0.150	0.150	0.150	0.150	0.109
KS Marsaglia Method	0.780	0.947	0.930	0.803	0.827	0.905	0.752	0.717	0.702	0.766	0.519
KS Lilliefors Modification	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.120
Anderson-Darling Test	0.384	0.564	0.687	0.654	0.764	0.730	0.582	0.757	0.674	0.612	0.246
Cramer- Von Mises Test	0.246	0.494	0.599	0.578	0.714	0.730	0.594	0.736	0.603	0.521	0.212
Shapiro-Wilk Test	0.781	0.844	0.898	0.899	0.930	0.932	0.842	0.873	0.876	0.850	0.606
Shapiro-Francia Test	0.475	0.666	0.772	0.758	0.811	0.799	0.737	0.825	0.841	0.797	0.558
Jarque-Bera Test	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
D'Agostino & Pearson Test	0.685	0.832	0.876	0.900	0.915	0.950	0.972	0.980	0.985	0.987	0.993

Based on data in Table 5, the Jarque-Bera Test and D'Agostino & Pearson test methods conclude that data do not have a normal distribution under conditions where the sample size is 500 or more for the significance level of α=0.05. Kolmogorov Smirnov, KS Stephens Modification, KS Marsaglia, KS Lilliefors Modification, Anderson-Darling Test and Cramer-Von Mises Test show that regardless of the number of samples, data have a normal distribution for the significance level of α=0.05. In conclusion, it can be said that the Jarque-Bera Test and D'Agostino & Pearson Test methods are affected most by the sample size when the coefficients of kurtosis and skewness are 0.25. At the same time, the Kolmogorov Smirnov, KS Stephens Modification, KS Marsaglia, KS Lilliefors Modification, Anderson-Darling Test, and Cramer-Von Mises methods are not affected by the sample size. In Table 6, the significance values obtained from the normality tests are given for 11 different sample sizes under conditions where the kurtosis and skewness coefficients are 0.50.

Table 5. *Normality test results for different samples in cases where kurtosis and skewness coefficients are (0.25, 0.25)*

Sample Size	10	20	30	40	50	100	200	300	400	500	900
Normality Tests	*p*	*p*	*p*	*p*	*p*	*p*	*p*	*p*	*p*	*p*	*p*
Kolmogorov Smirnov	0.732	0.792	0.803	0.936	0.965	0.932	0.936	0.957	0.957	0.931	0.862
KS Stephens Modification	0.150	0.150	0.150	0.150	0.150	0.150	0.150	0.150	0.150	0.150	0.150
KS Marsaglia Method	0.657	0.739	0.760	0.913	0.950	0.917	0.925	0.950	0.951	0.924	0.855
KS Lilliefors Modification	0.198	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.200
Anderson-Darling Test	0.208	0.246	0.410	0.545	0.684	0.581	0.464	0.576	0.506	0.484	0.136
Cramer- Von Mises Test	0.125	0.180	0.344	0.574	0.767	0.738	0.612	0.693	0.573	0.576	0.353
Shapiro-Wilk Test	0.541	0.547	0.672	0.732	0.767	0.666	0.332	0.230	0.143	0.081	0.005
Shapiro-Francia Test	0.262	0.342	0.477	0.528	0.569	0.455	0.242	0.184	0.119	0.066	0.006
Jarque-Bera Test	0.937	0.878	0.823	0.771	0.722	0.522	0.272	0.142	0.074	0.039	0.003
D'Agostino & Pearson Test	0.482	0.577	0.574	0.554	0.529	0.401	0.221	0.121	0.066	0.036	0.003

In Table 6, the Shapiro-Wilk Test, Shapiro-Francia Test, Jarque-Bera Test, and D'Agostino & Pearson Test methods prove that data do not have a normal distribution under the sample size of 200 or more for the significance level of α=0.05. The Kolmogorov Smirnov and KS Marsaglia methods, on the other hand, give the result that the data have a normal distribution for the significance level of α=0.05, regardless of the sample size. As a result, it can be said that the Shapiro-Wilk Test, Shapiro-Francia Test, Jarque-Bera Test, and D'Agostino & Pearson Test methods are affected most by the sample size when the coefficients of kurtosis and skewness are 0.50. At the same time, the Kolmogorov Smirnov and KS Marsaglia methods are not affected by the sample size.

Table 6. *Normality test results for different samples in cases where kurtosis and skewness coefficients are (0.50, 0.50)*

Sample Size	10	20	30	40	50	100	200	300	400	500	900
Normality Tests	*p*	*p*	*p*	*p*	*p*	*p*	*p*	*p*	*p*	*p*	*p*
Kolmogorov Smirnov	0.672	0.562	0.644	0.910	0.973	0.893	0.806	0.665	0.369	0.247	0.105
KS Stephens Modification	0.142	0.099	0.150	0.150	0.150	0.150	0.150	0.150	0.044	0.014	0.010
KS Marsaglia Method	0.596	0.507	0.597	0.882	0.961	0.873	0.789	0.649	0.358	0.239	0.102
KS Lilliefors Modification	0.148	0.104	0.171	0.200	0.200	0.200	0.200	0.200	0.042	0.014	0.001
Anderson-Darling Test	0.125	0.059	0.211	0.379	0.485	0.260	0.055	0.036	0.014	0.006	0.000
Cramer- Von Mises Test	0.074	0.062	0.191	0.474	0.663	0.442	0.187	0.098	0.035	0.014	0.001
Shapiro-Wilk Test	0.349	0.261	0.352	0.380	0.361	0.139	0.011	0.002	0.000	0.000	0.000
Shapiro-Francia Test	0.152	0.147	0.226	0.251	0.243	0.093	0.011	0.002	0.000	0.000	0.000
Jarque-Bera Test	0.771	0.594	0.458	0.353	0.272	0.074	0.005	0.000	0.000	0.000	0.000
D'Agostino & Pearson Test	0.271	0.288	0.243	0.199	0.161	0.054	0.006	0.001	0.000	0.000	0.000

4. CONCLUSION and DISCUSSION

This study aims to compare normality tests in terms of different sample sizes in data with normal distribution in terms of different kurtosis and skewness coefficients obtained simulatively. For this purpose, 55 data sets with different skewness and kurtosis coefficients and different sample sizes were produced simulatively. The analysis results according to 10 different normality tests showed that that normality tests are not affected by the sample size when the skewness and kurtosis coefficients are equal to or close to zero. However, in cases where the skewness and kurtosis coefficients moved away from zero, it was found that normality tests are affected by the sample size, and normality tests tend to give significant results, especially for n>200. It can also be said that the Kolmogorov Smirnov and KS Marsaglia tests are relatively less affected by sample size than other normality tests under all conditions. In other words, the Kolmogorov Smirnov and KS Marsaglia methods tend to accept the H_0 hypothesis. In studies conducted with data that did not show a normal distribution, it was seen that larger samples were needed for the Kolmogorov Smirnov method to reject the H_0 hypothesis; however, smaller samples were sufficient for the Anderson-Darling and Shapiro-Wilk methods (Ahad et al., 2011; Kundu et al., 2011). It can be said that the Kolmogorov Smirnov and KS Marsaglia methods tend to accept the H_0 hypothesis in normal and non-normal data. In cases where skewness and kurtosis coefficients are close to zero, researchers may be advised to use the Kolmogorov Smirnov and KS Marsaglia methods for small samples. However, instead of normality tests in large samples, histogram graphs or critical values as suggested by Kim (2013) and Mayer (2013) for z-scores may be used.

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors.

Orcid

Suleyman Demir https://orcid.org/0000-0003-3136-0423

REFERENCES

Abbott, M.L. (2011). *Understanding educational statistics using Microsoft Excel and SPSS*. Wiley & Sons, Inc.

Ahad, N.A., Yin, T.S., Othman, A.R., & Yaacob, C.R. (2011). Sensitivity of normality tests to non-normal data. *Sains Malaysiana, 40*(6), 637-641. https://core.ac.uk/download/pdf/11491563.pdf

Anderson, T.W., & Darling, D.A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *The Annals of Mathematical Statistics, 23*(2), 193-212. https://doi.org/10.1214/aoms/1177729437

Anderson, T.W., & Darling, D.A. (1954). A test of goodness of fit. *Journal of the American Statistical Association, 49*(268), 765-769. https://doi.org/10.1080/01621459.1954.10501232

Baykul, Y., & Güzeller, C.O. (2013). *Sosyal bilimler için istatistik: SPSS uygulamalı [Statistics for social sciences: SPSS applied]. Pegem Akademi.*

Bulmer, M.G. (1979). *Principles of Statistics.* Dover.

Byrne, B.M. (2010). *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming.* Taylor and Francis Group Publication.

Csörgö, S., & Faraway, J.J. (1996). The exact and asymptotic distributions of Cramer-von Mises statistics. *Journal of Royal Statistical Society. Series B (Methodological), 58*(1), 221-234.

D'Agostino, R.B., & Pearson, E.S. (1973). Tests for departures from normality. Empirical results for the distribution of b2 and √b1. *Biometrika, 60*(3), 613-622. https://doi.org/10.1093/biomet/60.3.613

Dellal, G.E., & Wilkinson, L. (1986). An analytic approximation to the distribution of Lilliefors's test statistic for normality. *The American Statistician, 40*(4), 294-296. https://doi.org/10.1080/00031305.1986.10475419

Demir, E., Saatcioğlu, Ö., & İmrol, F. (2016). Uluslararası dergilerde yayımlanan eğitim araştırmalarının normallik varsayımları açısından incelenmesi [Examination of educational researches published in international journals in terms of normality assumptions]. *Current Research in Education, 2*(3), 130-148.

Douglas G.B., & Edith, S. (2002). A test of normality with high uniform power. *Journal of Computational Statistics and Data Analysis 40*(3), 435-445. https://doi.org/10.1016/S0167-9473(02)00074-9

Facchinetti, S. (2009). A procedure to find exact critical values of Kolmogorov-Smirnov test. *Statistica Applicata – Italian Journal of Applied Statistics, 21*(3-4), 337-359.

Field, A. (2013). *Discovering statistics using SPSS.* Sage Publications.

Frain, J.C. (2007). Small sample power of tests of normality when the alternative is an α-stable distribution. *Trinity Economics Papers TEP-0207,* Trinity College Dublin, Department of Economics. http://www.tcd.ie/Economics/TEP/2007/TEP0207.pdf

George, D., & Mallery, M. (2010). *SPSS for Windows Step by Step: A Simple Guide and Reference, 17.0*. Pearson.

Gravetter, F., & Wallnau, L. (2014). *Essentials of statistics for the behavioral sciences.* Wadsworth.

Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2010). *Multivariate data analysis: A global perspective.* Prentice Hall.

Harter, H.L. (1961). Expected values of normal order statistics, *Biometrika, 48,* 151-65.

Howell, D.C. (2013). *Statistical methods for psychology*. Belmont, Wadsworth/Cengage Learning.

Jarque, C.M., & Bera, A.K. (1987). A test for normality of observations and regression residuals. *International Statistical Review, 55*(2), 163-172. https://doi.org/10.2307/1403192

Keskin, S. (2006). Comparison of several univariate normality tests regarding type I error rate and power of the test in simulation based small samples. *Journal of Applied Science Research 2*(5), 296-300.

Kim, H.Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restor Dent Endod, 38*(1), 52-4. https://doi.org/10.5395/rde.2013.38.1.52

Kline, R.B. (2011). *Methodology in the Social Sciences: Principles and practice of structural equation modeling.* Guilford Press.

Kundu, M.G., Mishra, S., & Khare, D. (2011). Specificity and Sensitivity of Normality Tests. *In Proceedings of VI International Symposium on Optimisation and Statistics.* Anamaya Publisher.

Lee, C., Park, S., & Jeong, J. (2016). Comprehensive Comparison of Normality Tests: Empirical Study Using Many Different Types of Data. *Journal of the Korean Data and Information Science Society, 27*(5), 1399-1412. https://doi.org/10.7465/jkdi.2016.27.5.1399

Lilliefors, H.W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association, 62*(318), 399-402. https://doi.org/10.1080/01621459.1967.10482916

Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The importance of the normality assumption in large public health data sets. *Annual Review of Public Health, 23,* 151–169. https://doi.org/10.1146/annurev.publhealth.23.100901.140546

Marsaglia, G., Tsang, W.W., & Wang, J. (2003). Evaluating Kolmogorov's distribution. *Journal of Statistical Software, 8*(18). https://doi.org/10.18637/jss.v008.i18

Martin, W., & Bridgmon, K. (2012). *Quantitative and statistical research methods: from hypothesis to results.* Jossey-Bass.

Mayers, A. (2013). *Introduction to statistics and SPSS in psychology.* Pearson Education Limited.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105*(1), 156–166. https://doi.org/10.1037/0033-2909.105.1.156

Nor-Aishah H., & Shamsul R. A (2007, 12-14 December). *Robust Jacque-Bera Test of Normality.* The 9th Islamic Countries Conference on Statistical Sciences, University Malaya, Malaysia.

Nornadiah, M.R., & Yap, B.W. (2011). Power comparison of Shapiro-Wilk, Kolmogorov-Smirnov, Lillieforsand Anderson-Darling tests. *Journal of Statistical Modeling and Analytics, 2*(1), 21-33.

Orcan, F. (2020). Parametric or non-parametric: skewness to test normality for mean comparison. *International Journal of Assessment Tools in Education, 7*(2), 255-265. https://doi.org/10.21449/ijate.656077

Öztuna, D., Elhan, A.H., & Tuccar, E. (2006). Investigation of four different normality tests in terms of Type I error rate and power under different distributions, *Turk. J. Med. Sci. 36*(3), 171–176.

Öner, M., & Kocakoç, İ.D. (2017). JMASM 49: A compilation of some popular goodness of fit tests for normal distribution: their algorithms and MATLAB codes (MATLAB). *Journal of Modern Applied Statistical Methods, 16*(2), 547-575. https://doi.org/10.22237/jmasm/1509496200

Rinnakorn, C., & Kamon, B. (2007). A power comparison of goodness-of-fit tests for normality based on the likelihood ratio and the non-likelihood ratio. *Thailand Statistician, 5,* 57-68.

Shapiro, S.S., & Francia, R.S. (1972). An approximate analysis of variance test for normality. *Journal of the American Statistical Association, 67*(337), 215-216. https://doi.org/10.1080/01621459.1972.10481232

Stephens, M.A. (1986). Tests based on EDF statistics. In R. B. D'Agostino & M. A. Stephens (Eds.), *Goodness-of-fit techniques* (pp. 97-194). Marcel Dekker.

Stephens, M.A. (1974), EDF Statistics for Goodness of Fit and Some Comparisons. *Journal of the American Statistical Association, 69*, 730-737. https://doi.org/10.2307/2286009

Shapiro, S.S., & Wilk, M.B. (1965). An analysis of variance test for normality. *Biometrika, 52,* 591-611. https://doi.org/10.2307/2333709

Smirnov, N.V. (1948). Table for estimating the goodness of t of empirical distributions. *The Annals of Mathematical Statistics, 19,* 279-281.

Tabachnick, B.G., & Fidell, L.S. (2013). *Using Multivariate Statistics.* Pearson.

Trochim, W.M., & Donnelly, J.P. (2006). *The research methods knowledge base.* Atomic Dog.

Ukponmwan, H.N., & Ajibade, F.B. (2017). Evaluation of techniques for univariate normality test using monte carlo simulation. *American Journal of Theoretical and Applied Statistics*, *6*(5), 51-61. https://doi.org/10.11648/j.ajtas.s.2017060501.18

Wilcox, R.R. (2010). *Fundamentals of modern statistical methods: Substantially improving power and accuracy.* Springer-Verlag.

Yap, B.W., & Sim, C.H. (2011). Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation, 81*(12), 2141-2155. https://doi.org/10.1080/00949655.2010.520163

APPENDIX

Description of Normality Tests

Kolmogorov Smirnov Test	$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{X_i \le x}$ $D_n = \max_x \lvert F_n(x) - F(x) \rvert$	$F_n(x)$ = the empirical distribution function n= number of observations $I_{X_i \le x}$ = the indicator function If the value of D_n is greater than the critical value, the null hypothesis is rejected.
KS Stephens Modification	$D^* = D_n\left(\sqrt{n} - 0.01 + \frac{0.85}{\sqrt{n}}\right)$ $p-value = \begin{cases} D^* < 0.775 & p > 0.15 \\ 0.775 < D^* < 0.819 & 0.10 < p < 0.15 \\ 0.819 < D^* < 0.895 & 0.05 < p < 0.10 \\ 0.895 < D^* < 0.995 & 0.025 < p < 0.05 \\ 0.995 < D^* < 1.035 & 0.01 < p < 0.025 \\ D^* \ge 1.035 & p < 0.01 \end{cases}$	n= number of observations If p-value is smaller than 0.05, the null hypothesis is rejected.

KS Marsaglia Method	$\Pr(D_n \le d) = \frac{n!}{n^n} H^n$, $d = \frac{k-h}{n}$ $p-value = \Pr(D_n > d) = 1 - \Pr(D_n \le d) = 1 - \frac{n!}{n^n} H^n$	n= number of observations H is a *m x m* matrix that depends on h only. k is a positive integer $0 \le h < 1$ If p-value is smaller than 0.05, the null hypothesis is rejected.
KS Lilliefors Modification	$p-value = \begin{cases} D_n = D10 & 0.10 \\ D_n > D10 & \exp\left(aD_n^{\ 2} + bD_n + c - 2.3025851\right) \\ D_n \ge D15 & 0.15 + \left(D_n - D15\right)\left[\left(0.10 - 0.15\right)/\left(D10 - D15\right)\right] \\ D_n \ge D20 & 0.20 + \left(D_n - D20\right)\left[\left(0.15 - 0.20\right)/\left(D15 - D20\right)\right] \\ D_n \le D10 & p > 0.20 \end{cases}$	D10, D15 and D20 corresponding to n from the table given by Dellal and Wilkinson (1986). If p-value is smaller than 0.05, the null hypothesis is rejected.