Communications for Statistical Applications and Methods
2015, Vol. 22, No. 4, 361–375

DOI: http://dx.doi.org/10.5351/CSAM.2015.22.4.361
Print ISSN 2287-7843 / Online ISSN 2383-4757

Tests Based on Skewness and Kurtosis for Multivariate Normality

Namhyun Kim[1,a]

[a]Department of Science, Hongik University, Korea

Abstract

A measure of skewness and kurtosis is proposed to test multivariate normality. It is based on an empirical standardization using the scaled residuals of the observations. First, we consider the statistics that take the skewness or the kurtosis for each coordinate of the scaled residuals. The null distributions of the statistics converge very slowly to the asymptotic distributions; therefore, we apply a transformation of the skewness or the kurtosis to univariate normality for each coordinate. Size and power are investigated through simulation; consequently, the null distributions of the statistics from the transformed ones are quite well approximated to asymptotic distributions. A simulation study also shows that the combined statistics of skewness and kurtosis have moderate sensitivity of all alternatives under study, and they might be candidates for an omnibus test.

Keywords: Goodness of fit tests, multivariate normality, skewness, kurtosis, scaled residuals, empirical standardization, power comparison

1. Introduction

Classical multivariate analysis techniques require the assumption of multivariate normality; consequently, there are numerous test procedures to assess the assumption in the literature. For a general review, some references are Henze (2002), Henze and Zirkler (1990), Srivastava and Mudholkar (2003), and Thode (2002, Ch.9). For comparative studies in power, we refer to Farrell *et al.* (2007), Horswell and Looney (1992), Mecklin and Mundfrom (2005), and Romeu and Ozturk (1993).

Most multivariate techniques are often generalizations of univariate normality tests. W test by Shapiro and Wilk (1965) and moment tests based on skewness and kurtosis are some of the most popular tests for univariate normality (Pearson *et al.*, 1977). Shapiro and Wilk's W test and the approximate tests to that (De Wet and Venter, 1972; Shapiro and Francia, 1972) have been generalized to multivariate cases by Fattorini (1986), Kim (2004a, 2005), Kim and Bickel (2003), Malkovich and Afifi (1973), Mudholkar *et al.* (1995), Royston (1983), Srivastava and Hui (1987), and Villasenor Alva and González Estrada (2009). As for skewness and kurtosis approaches, Mardia (1970, 1974)'s procedures are the most frequently used tests for multivariate normality. Kim (2004b), and Malkovich and Afifi (1973) generalized univariate moments to multivariate ones based on linear combinations of variates. Srivastava (1984) used principal components.

Numerous different testing procedures have been proposed in the literature to test multinormality; however, research has indicated no single test to be the most powerful for all situations. The Henze

This work was supported by 2015 Hongik University Research Fund. This work was done while the author was on sabbatical (Feb. 2015-Jan. 2016).
[1] Department of Science, Hongik University, 94 Wausan-Ro, Mapo-Gu, Seoul 121-791, Korea. E-mail: nhkim@hongik.ac.kr

Published 31 July 2015 / journal homepage: http://csam.or.kr

and Zirkler (1990) test is usually recommended as a formal goodness of fit test for multinormality since the invariance and consistency are proven theoretically and it has relatively good power across a wide range of alternatives (Farrell *et al.*, 2007; Mecklin and Mundfrom, 2005). However some supplementary or less formal procedures (such as Mardia's skewness and kurtosis measures or some graphical procedures) should be followed up to help diagnose possible deviation from normality.

Mardia's procedures are among the most commonly used tests for multinormality; however, Baringhaus and Henze (1992), and Henze (1994) showed that Mardia's skewness and kurtosis are inconsistent against certain alternatives. Consequently, Mardia's procedures may have low power against some alternatives. It could also happen for other procedures for other procedures based on skewness or kurtosis. However, skewness and kurtosis can provide direct measure of departure from normality with strong points over other procedures.

In this paper, we propose multivariate skewness and kurtosis statistics based on an empirical standardization using the scaled residuals of the observations. Villasenor Alva and González Estrada (2009) used this idea to generalize Shapiro and Wilk's W test that is closely related to Srivastava and Hui (1987)'s principal component approach. Section 2 presents the proposed test statistics for multivariate normality with the simulated critical values and p-values using asymptotic distributions. Section 3 contains a real data example and simulation study to compare the power performance of the statistics. Section 4 ends the paper with the concluding remarks.

2. Test Statistics Based on Skewness and Kurtosis

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be p-dimensional independent and identically distributed (i.i.d.) random vectors, and let $N_p(\mu, \mathbf{\Sigma})$ be the p-variate multivariate normal distribution with mean vector μ and covariance matrix $\mathbf{\Sigma}$. For the univariate normal distribution with mean μ and variance σ^2, we will use $N(\mu, \sigma^2)$ omitting the index.

We want to test the null hypothesis

$$H_0 : \mathbf{X}_1, \ldots, \mathbf{X}_n \text{ is a sample from } N_p(\mu, \mathbf{\Sigma}) \text{ for some } \mu \text{ and } \mathbf{\Sigma}.$$

First, we transform $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ as

$$\mathbf{Z}_i = \mathbf{S}^{*'} \left(\mathbf{X}_i - \bar{\mathbf{X}}\right), \quad i = 1, \ldots, n, \tag{2.1}$$

where ' $'$ ' denotes a transpose, $\bar{\mathbf{X}}$ is the sample mean vector, $\bar{\mathbf{X}} = n^{-1} \sum_{j=1}^{n} \mathbf{X}_j$, $\mathbf{S}$ is the sample covariance matrix, $\mathbf{S} = n^{-1} \sum_{j=1}^{n} (\mathbf{X}_j - \bar{\mathbf{X}})(\mathbf{X}_j - \bar{\mathbf{X}})'$, and $\mathbf{S}^*$ is defined by $\mathbf{S}^{*'}\mathbf{S}\mathbf{S}^* = \mathbf{I}$. We call $\mathbf{Z}_i$'s the scaled residuals, and it is well known $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ follow $N_p(\mathbf{0}, \mathbf{I})$ asymptotically if $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ is a sample from $N_p(\mu, \mathbf{\Sigma})$, where $\mathbf{0}$ is the null vector of order p and $\mathbf{I}$ is the identity matrix of order $p \times p$. The components of $\mathbf{Z}_i$, denoted by $(Z_{1i}, Z_{2i}, \ldots, Z_{pi})$, are approximately independent standard normal $N(0, 1)$ under the null hypothesis.

Therefore we can think of test statistics that consider the skewness or the kurtosis of each coordinate $(Z_{k1}, Z_{k2}, \ldots, Z_{kn})$, $k = 1, \ldots, p$. As for skewness, it is defined by

$$\sqrt{b_1(k)} = \frac{\sqrt{n} \sum_{j=1}^{n} \left(Z_{kj} - \bar{Z}_k\right)^3}{\left[\sum_{j=1}^{n} \left(Z_{kj} - \bar{Z}_k\right)^2\right]^{\frac{3}{2}}} = \frac{1}{n} \sum_{j=1}^{n} Z_{kj}^3, \tag{2.2}$$

and it is usually a two-sided test. Its square

$$b_1(k) = \frac{n\left[\sum_{j=1}^{n}\left(Z_{kj} - \bar{Z}_k\right)^3\right]^2}{\left[\sum_{j=1}^{n}\left(Z_{kj} - \bar{Z}_k\right)^2\right]^3} = \left(\frac{1}{n}\sum_{j=1}^{n} Z_{kj}^3\right)^2 \tag{2.3}$$

is a one-sided test. The second equality in (2.2) or (2.3) follows from that each coordinate $(Z_{k1}, Z_{k2}, \ldots, Z_{kn})$, $k = 1, \ldots, p$ has mean 0 and variance 1.

To test the null hypothesis, we can think of the test statistic

$$B_1 = \max_{1 \le k \le p} b_1(k), \tag{2.4}$$

since large value of $b_1(k)$ indicates a departure from normality. The statistic in (2.4) is closely related to Srivastava (1984)'s. Malkovich and Afifi (1973) generalized the univariate skewness and kurtosis to test multivariate normality using Roy's union-intersection principle (Roy, 1953), which is based on the fact that $\mathbf{c}'\mathbf{X}$ follows a univariate normal for all $\mathbf{c}$, $\mathbf{c} \neq \mathbf{0}$, if $\mathbf{X}$ follows a multivariate normal. They investigated the skewness or the kurtosis of all the possible linear combinations that reduce to normal under the null hypothesis and tried to find a direction that gives furthest away from normal. Malkovich and Afifi (1973)'s multivariate skewness is as follows.

$$b_{1,M} = \max_{\boldsymbol{c},\ \boldsymbol{c} \neq \boldsymbol{0}} b_1(\boldsymbol{c}) = \max_{\boldsymbol{c},\ \|\boldsymbol{c}\|=1} \frac{1}{n^2}\left[\sum_{j=1}^{n}\left(\boldsymbol{c}'\boldsymbol{Z}_j\right)^3\right]^2, \tag{2.5}$$

where $\sqrt{b_1(\mathbf{c})}$ is the univariate skewness of $\boldsymbol{c}'\boldsymbol{X}_1, \ldots, \boldsymbol{c}'\boldsymbol{X}_n$ as in (2.2) by substituting $y_j = \boldsymbol{c}'\boldsymbol{X}_j$ into Z_{kj}, $b_1(\mathbf{c})$ is its square likewise, and $\boldsymbol{Z}_j$ is the scaled residual in (2.1). The second equality in (2.5) comes from the affine invariance with respect to addition of vectors and multiplication of nonsingular matrices (Kim, 2004b). However the statistic in (2.5) is hard to compute, especially when the dimension p becomes large, practically when the dimension is $p > 2$ (Horswell and Looney, 1992). Hence Kim (2004b) proposed an approximation to Malkovich and Afifi's statistic by selecting some directions that might achieve the maximum in (2.5) asymptotically. The statistic is included for comparison in Section 3. However, B_1 in (2.4) selects the vectors $\mathbf{c}$ as the unit vectors in each coordinate such that $\mathbf{c} = (1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)$.

Under the null hypothesis of normality, $\sqrt{b_1(k)}$ is asymptotically normal with mean 0 and variance $6/n$ (Kendall and Stuart, 1977). Specifically,

$$\sqrt{b_1(k)}\left[\frac{(n+1)(n+3)}{6(n-2)}\right]^{\frac{1}{2}} \xrightarrow{d} N(0, 1). \tag{2.6}$$

Therefore the limit distribution of B_1 in (2.4) becomes

$$B_1^* := \frac{(n+1)(n+3)}{6(n-2)} B_1 \xrightarrow{d} \max_{1 \le k \le p} V_k, \tag{2.7}$$

where $V_1, \ldots, V_p$ follow i.i.d. χ_1^2, chi-square distribution with 1 degree of freedom, and we have

$$\lim_{n \to \infty} P\left(B_1^* \le x\right) = \left(P(V_1 \le x)\right)^p.$$

However, the sample size must be somewhat large so that the normal approximation in (2.6) can be used. According to D'Agostino (1986), it appears to be valid for $n \geq 150$.

D'Agostino (1970) presented a transformation of the null distribution of $\sqrt{b_1(k)}$ to normality using a Johnson's unbounded S_U curve. For each $\sqrt{b_1(k)}$, we compute

$$\begin{aligned}
Y_k^* &= \sqrt{b_1(k)}\left[\frac{(n+1)(n+3)}{6(n-2)}\right]^{\frac{1}{2}}, \\
\beta_2 &= \frac{3\left(n^2+27n-70\right)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)}, \\
\omega^2 &= \sqrt{2(\beta_2-1)}-1, \\
\delta &= \frac{1}{\sqrt{\log\omega}}, \\
\alpha &= \sqrt{\frac{2}{\omega^2-1}},
\end{aligned}$$

and

$$T_k^* := T_k^*\left(\sqrt{b_1(k)}\right) = \delta\log\left(\frac{Y_k^*}{\alpha}+\sqrt{\left(\frac{Y_k^*}{\alpha}\right)^2+1}\right). \tag{2.8}$$

The constant β_2 in the above transformation is based on the fourth standardized moment of the distribution of the skewness. The T_k^*'s follow approximately a standard normal $N(0,1)$. It is applicable even for small sample sizes, $n \geq 8$. We propose the following statistics using T_k^* in (2.8)

$$S_1 = \max_{1\leq k\leq p}\left(T_k^*\right)^2, \qquad S_2 = \sum_{k=1}^{p}\left(T_k^*\right)^2$$

as test statistics to test multivariate normality, then we have

$$S_1 \xrightarrow{d} \max_{1\leq k\leq p} V_k, \quad \text{and} \quad S_2 \xrightarrow{d} \chi_p^2.$$

Likewise B_1 in (2.4), large values of S_1 or S_2 will indicate non-normality.

For kurtosis, it is obtained by

$$b_2(k) = \frac{n\sum_{j=1}^{n}\left(Z_{kj}-\bar{Z}_k\right)^4}{\left[\sum_{j=1}^{n}\left(Z_{kj}-\bar{Z}_k\right)^2\right]^2} = \frac{1}{n}\sum_{j=1}^{n} Z_{kj}^4, \tag{2.9}$$

and the kurtosis is asymptotically normal with mean 3 and variance $24/n$ under normality. The following normal approximation is often used

$$Y_k^{**} := \sqrt{\frac{(n+1)^2(n+3)(n+5)}{24n(n-2)(n-3)}}\left(b_2(k)-\frac{3(n-1)}{(n+1)}\right) \xrightarrow{d} N(0,1) \tag{2.10}$$

using the mean and variance of the kurtosis. A test statistic for multivariate normality could be

$$B_2 = \max_{1\le k\le p}\left(b_2(k) - \frac{3(n-1)}{(n+1)}\right)^2,$$

and we have

$$B_2^{**} := \frac{(n+1)^2(n+3)(n+5)}{24n(n-2)(n-3)} B_2 \xrightarrow{d} \max_{1\le k\le p} V_k. \tag{2.11}$$

The normal approximation for the kurtosis in (2.10) is slow and the sample size must be extremely large so that the approximation is valid. D'Agostino (1986) recommended it not be used because it should be over $n = 1000$.

Likewise in the skewness, Malkovich and Afifi (1973)'s multivariate kurtosis is defined by

$$b_{2,M}^2 = \max_{\boldsymbol{c},\boldsymbol{c}\neq\boldsymbol{0}}[b_2(\boldsymbol{c}) - 3]^2 = \max_{\boldsymbol{c},\,\|\boldsymbol{c}\|=1}\left[\frac{1}{n}\sum_{j=1}^{n}\left(\boldsymbol{c}'\boldsymbol{Z}_j\right)^4 - 3\right]^2 \tag{2.12}$$

with $b_2(\mathbf{c})$ the univariate kurtosis of $\boldsymbol{c}'\boldsymbol{X}_1, \ldots, \boldsymbol{c}'\boldsymbol{X}_n$ as in (2.9) plugging $y_j = \boldsymbol{c}'\boldsymbol{X}_j$ into Z_{kj}. The approximate statistic is given in Kim (2004b).

Anscombe and Glynn (1983) showed a normal approximation for the kurtosis. Compute $\sqrt{\beta_1}$, the third standardized moment of kurtosis,

$$\sqrt{\beta_1} = \frac{6(n^2-5n+2)}{(n+7)(n+9)}\sqrt{\frac{6(n+3)(n+5)}{n(n-3)(n-2)}},$$

and compute

$$A = 6 + \frac{8}{\sqrt{\beta_1}}\left[\frac{2}{\sqrt{\beta_1}} + \sqrt{1 + \frac{4}{\beta_1}}\right].$$

Finally

$$T_k^{**} := T_k^{**}(b_2(k)) = \left[1 - \frac{2}{9A} - \left(\frac{1-(2/A)}{1 + Y_k^{**}\sqrt{2/(A-4)}}\right)^{\frac{1}{3}}\right]\frac{1}{\sqrt{2/(9A)}} \tag{2.13}$$

follows approximately a standard normal $N(0, 1)$. The Y_k^{**} is defined in (2.10), and it is the standardized value of the kurtosis. The proposed statistic using T_k^{**} in (2.13) could be

$$K_1 = \max_{1\le k\le p}(T_k^{**})^2, \qquad K_2 = \sum_{k=1}^{p}(T_k^{**})^2$$

likewise in the skewness. They have

$$K_1 \xrightarrow{d} \max_{1\le k\le p} V_k, \quad \text{and} \quad K_2 \xrightarrow{d} \chi_p^2.$$

Table 1: Simulated critical values and p-values using asymptotic distribution for $p = 2$, $\alpha = 0.05$

Statistic	Values	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$	$n = 100$	$n = 150$	$n = 200$	$n = \infty$
B_1	Critical value	1.811	1.201	0.882	0.688	0.563	0.289	0.203	0.151	
B_1^*	Critical value	5.396	5.371	5.373	5.324	5.284	5.107	5.286	5.173	5.002
	p-value	0.040	0.041	0.040	0.042	0.043	0.047	0.043	0.045	0.050
S_1	Critical value	5.317	5.057	5.010	4.855	5.044	5.007	4.978	5.095	5.002
	p-value	0.042	0.048	0.050	0.054	0.049	0.050	0.051	0.047	0.050
S_2	Critical value	5.908	5.983	6.175	5.809	5.933	5.940	5.910	5.927	5.991
	p-value	0.052	0.050	0.046	0.055	0.051	0.051	0.052	0.052	0.050
B_2	Critical value	3.687	3.628	3.124	2.805	2.239	1.110	0.836	0.599	
B_2^{**}	Critical value	6.472	6.263	6.370	6.761	6.270	5.368	5.770	5.382	5.002
	p-value	0.022	0.025	0.023	0.019	0.024	0.041	0.032	0.040	0.050
K_1	Critical value	4.479	5.027	4.913	4.917	5.219	5.059	5.080	5.238	5.002
	p-value	0.067	0.049	0.053	0.052	0.044	0.048	0.048	0.044	0.050
K_2	Critical value	5.406	5.976	5.893	6.074	6.073	6.359	6.321	6.075	5.991
	p-value	0.067	0.050	0.053	0.048	0.048	0.042	0.042	0.048	0.050
C_1	Critical value	5.284	5.658	5.997	6.069	6.261	6.317	6.352	6.171	6.205
	p-value	0.083	0.068	0.056	0.054	0.048	0.047	0.046	0.051	0.050
C_2	Critical value	10.627	10.581	10.172	10.200	10.441	10.368	10.168	9.849	9.488
	p-value	0.031	0.032	0.038	0.037	0.034	0.035	0.038	0.043	0.050

Srivastava and Hui (1987) computed Shapiro and Wilk's W statistics of the principal components, transformed them to normality, applied the transformation $-2\ln\phi(\cdot)$ for them with ϕ the density of a standard normal and took a summation for asymptotic distribution χ^2_{2p}. When we apply $-2\ln\phi(\cdot)$ to the transformed skewness T_k^* or the transformed kurtosis T_k^{**} and make the statistic $-2\sum_{k=1}^p \ln\phi(T_k^*)$ or $-2\sum_{k=1}^p \ln\phi(T_k^{**})$, these statistics should not be used due to their extremely poor power.

Considerable attention has been paid to the omnibus tests that combine information from skewness and kurtosis. D'Agostino and Pearson (1973) proposed the statistic

$$\left(T_k^*\left(\sqrt{b_1(k)}\right)\right)^2 + \left(T_k^{**}\left(b_2(k)\right)\right)^2$$

when $k = 1$ as an omnibus statistic for testing univariate normality and viewed it as χ^2_2 although the two although two variables are not independent. D'Agostino and Pearson (1974), D'Agostino (1986), and Bowman and Shenton (1986) mentioned that they are uncorrelated and nearly independent, and χ^2_2 approximation for the statistic is no problem for $n \geq 100$. Bowman and Shenton (1975, 1986) also gave the same format of the statistic; therefore, we may propose the following statistics

$$C_1 = \max_{1\leq k\leq p}\left[\left(T_k^*\right)^2, \left(T_k^{**}\right)^2\right], \qquad C_2 = \sum_{k=1}^{p}\left(T_k^*\right)^2 + \sum_{k=1}^{p}\left(T_k^{**}\right)^2$$

as p-dimensional omnibus test statistics, having the null distributions approximately $\max_{1\leq k\leq 2p} V_k$, χ^2_{2p} respectively, with $V_1, \ldots, V_{2p}$ i.i.d. χ^2_1.

A simulation study was performed to study the null distributions of the proposed statistics. For each combination of dimensions $p = 2, 5$ and different sample sizes $n = 10, 20, 30, 40, 50, 100, 150, 200$, and for $p = 10$, $n = 30, 40, 50, 100, 150, 200, 250, 300$, we generate $N = 5{,}000$ random samples and calculate the statistics. Random samples are generated from independent $N_p(\mathbf{0}, \mathbf{I})$, since the distribution of the statistics do not depend on unknown parameters $\mu, \mathbf{\Sigma}$.

In Table 1–6, the simulated critical values and p-values for each statistic are given for $\alpha = 0.05, 0.10$. The simulated quantile values k_α are given in the tables. The p-values are obtained using the asymptotic distribution $\max_{1\leq k\leq p} V_k$ for B_1, B_2, S_1, K_1, $\max_{1\leq k\leq 2p} V_k$ for C_1, χ^2_p for S_2, K_2, and χ^2_{2p} for C_2.

Table 2: Simulated critical values and p-values using asymptotic distribution for $p = 2$, $\alpha = 0.10$

Statistic	Values	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$	$n = 100$	$n = 150$	$n = 200$	$n = \infty$
B_1	Critical value	1.334	0.867	0.625	0.493	0.418	0.216	0.151	0.115	
B_1^*	Critical value	3.975	3.876	3.808	3.813	3.923	3.824	3.919	3.938	3.798
	p-value	0.090	0.096	0.099	0.099	0.093	0.099	0.093	0.092	0.100
S_1	Critical value	3.883	3.843	3.851	3.814	3.790	3.839	3.850	3.866	3.798
	p-value	0.095	0.097	0.097	0.099	0.100	0.098	0.097	0.096	0.100
S_2	Critical value	4.521	4.529	4.645	4.572	4.613	4.562	4.691	4.537	4.605
	p-value	0.104	0.104	0.098	0.102	0.100	0.102	0.096	0.103	0.100
B_2	Critical value	2.150	1.957	1.736	1.556	1.224	0.712	0.507	0.402	
B_2^{**}	Critical value	3.775	3.378	3.54	3.749	3.429	3.443	3.498	3.608	3.798
	p-value	0.101	0.128	0.116	0.103	0.124	0.123	0.119	0.112	0.100
K_1	Critical value	3.506	3.785	3.823	3.790	3.907	3.844	3.874	3.934	3.798
	p-value	0.119	0.101	0.099	0.100	0.094	0.097	0.096	0.092	0.100
K_2	Critical value	4.228	4.562	4.511	4.713	4.615	4.727	4.825	4.602	4.605
	p-value	0.121	0.102	0.105	0.095	0.100	0.094	0.090	0.100	0.100
C_1	Critical value	4.184	4.456	4.691	4.765	5.030	5.007	5.060	4.881	4.956
	p-value	0.154	0.132	0.116	0.111	0.096	0.097	0.094	0.104	0.100
C_2	Critical value	7.934	7.886	7.936	7.797	8.290	8.073	8.092	7.817	7.779
	p-value	0.094	0.096	0.094	0.099	0.082	0.089	0.088	0.099	0.100

Table 3: Simulated critical values and p-values using asymptotic distribution for $p = 5$, $\alpha = 0.05$

Statistic	Values	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$	$n = 100$	$n = 150$	$n = 200$	$n = \infty$
B_1	Critical value	2.530	1.702	1.222	0.995	0.788	0.406	0.267	0.194	
B_1^*	Critical value	7.538	7.614	7.442	7.691	7.397	7.180	6.936	6.677	6.599
	p-value	0.030	0.029	0.031	0.027	0.032	0.036	0.042	0.048	0.050
S_1	Critical value	6.642	6.707	6.634	6.326	6.813	6.631	6.520	6.464	6.599
	p-value	0.049	0.047	0.049	0.058	0.044	0.049	0.052	0.054	0.050
S_2	Critical value	11.278	10.910	11.255	11.245	11.065	11.141	11.228	11.118	11.070
	p-value	0.046	0.053	0.047	0.047	0.050	0.049	0.047	0.049	0.050
B_2	critical value	6.545	6.805	5.514	4.642	4.025	2.198	1.306	0.970	
B_2^{**}	Critical value	11.491	11.748	11.244	11.187	11.273	10.628	9.019	8.713	6.599
	p-value	0.003	0.003	0.004	0.004	0.004	0.006	0.013	0.016	0.050
K_1	Critical value	5.805	6.422	6.689	6.958	6.871	6.991	7.179	7.252	6.599
	p-value	0.077	0.055	0.048	0.041	0.043	0.040	0.036	0.035	0.050
K_2	Critical value	10.033	10.552	11.096	11.483	11.343	11.550	11.616	11.600	11.070
	p-value	0.074	0.061	0.050	0.043	0.045	0.041	0.040	0.041	0.050
C_1	Critical value	6.783	7.279	7.495	7.710	7.830	8.063	8.134	8.236	7.838
	p-value	0.088	0.068	0.060	0.054	0.050	0.044	0.043	0.040	0.050
C_2	Critical value	19.892	20.087	20.001	19.697	19.478	19.056	19.539	18.757	18.307
	p-value	0.030	0.028	0.029	0.032	0.035	0.040	0.034	0.043	0.050

They are $P(\max_{1\le k\le p} V_k \ge k_\alpha)$, $P(\max_{1\le k\le 2p} V_k \ge k_\alpha)$, $P(\chi^2_p \ge k_\alpha)$, or $P(\chi^2_{2p} \ge k_\alpha)$ for each case. As for B_1, B_2, the statistics B_1^* in (2.7), B_2^{**} in (2.11) are the ones where suitable constants depending on n are multiplied with corresponding asymptotic distributions. For skewness, the p-values are about right even for B_1 when $p = 2$, however they become different from the given significance level α when $p = 5, 10$. Regarding S_1, S_2, the null distributions are well approximated by their asymptotic distributions $\max_{1\le k\le p} V_k, \chi^2_p$, respectively even for small sample sizes. As for kurtosis, the p-values for B_2 show a considerable difference from the given level and are significantly worse for $p = 5, 10$. The null distribution of B_2 is extremely positively skewed, and asymptotic distribution should not be used. For K_1, K_2, and C_1, C_2, the sample size n should be moderately large to provide the right significance level for $p = 2, 5$, and they are conservative for $p = 10$. The asymptotic distributions

Table 4: Simulated critical values and p-values using asymptotic distribution for $p = 5$, $\alpha = 0.10$

Statistic	Values	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$	$n = 100$	$n = 150$	$n = 200$	$n = \infty$
B_1	Critical value	1.956	1.336	0.948	0.760	0.612	0.313	0.215	0.160	
B_1^*	Critical value	5.828	5.974	5.773	5.874	5.744	5.540	5.593	5.499	5.339
	p-value	0.076	0.070	0.079	0.075	0.080	0.090	0.087	0.092	0.100
S_1	Critical value	5.468	5.380	5.277	5.189	5.364	5.280	5.397	5.351	5.339
	p-value	0.093	0.098	0.103	0.109	0.099	0.103	0.097	0.099	0.100
S_2	Critical value	9.447	9.261	9.425	9.407	9.181	9.301	9.389	9.295	9.236
	p-value	0.092	0.099	0.093	0.094	0.102	0.098	0.095	0.098	0.100
B_2	Critical value	4.210	4.340	3.293	2.984	2.639	1.333	0.867	0.671	
B_2^{**}	Critical value	7.391	7.492	6.714	7.191	7.392	6.446	5.984	6.029	5.339
	p-value	0.032	0.031	0.047	0.036	0.032	0.054	0.070	0.068	0.100
K_1	Critical value	4.731	5.251	5.331	5.648	5.440	5.587	5.602	5.542	5.339
	p-value	0.140	0.105	0.100	0.084	0.095	0.087	0.087	0.089	0.100
K_2	Critical value	8.525	8.767	9.255	9.444	9.337	9.635	9.576	9.575	9.236
	p-value	0.130	0.119	0.099	0.093	0.096	0.086	0.088	0.088	0.100
C_1	Critical value	5.755	6.009	6.249	6.374	6.539	6.707	6.650	6.833	6.551
	p-value	0.153	0.134	0.118	0.110	0.101	0.092	0.095	0.086	0.100
C_2	Critical value	16.604	16.944	16.863	16.731	16.854	16.285	16.541	16.244	15.987
	p-value	0.084	0.076	0.077	0.081	0.078	0.092	0.085	0.093	0.100

Table 5: Simulated critical values and p-values using asymptotic distribution for $p = 10$, $\alpha = 0.05$

Statistic	values	$n = 30$	$n = 40$	$n = 50$	$n = 100$	$n = 150$	$n = 200$	$n = 250$	$n = 300$	$n = \infty$
B_1	Critical value	1.523	1.177	0.964	0.492	0.326	0.246	0.192	0.160	
B_1^*	Critical value	9.274	9.105	9.044	8.696	8.476	8.443	8.204	8.136	7.838
	p-value	0.023	0.025	0.026	0.031	0.035	0.036	0.041	0.043	0.050
S_1	Critical value	7.823	7.826	7.732	7.968	8.056	7.814	7.726	7.720	7.838
	p-value	0.050	0.050	0.053	0.047	0.044	0.051	0.053	0.053	0.050
S_2	Critical value	18.405	18.496	17.878	17.981	18.421	18.230	18.455	18.496	18.307
	p-value	0.049	0.047	0.057	0.055	0.048	0.051	0.048	0.047	0.050
B_2	Critical value	8.378	7.063	5.959	3.005	1.926	1.331	1.077	0.843	
B_2^{**}	Critical value	17.084	17.022	16.689	14.533	13.299	11.954	11.906	11.073	7.838
	p-value	0.000	0.000	0.000	0.001	0.003	0.005	0.006	0.009	0.050
K_1	Critical value	7.974	8.269	8.527	8.777	8.676	8.702	8.472	8.557	7.838
	p-value	0.046	0.040	0.034	0.030	0.032	0.031	0.035	0.034	0.050
K_2	Critical value	18.602	19.047	19.012	19.296	19.273	18.889	18.711	18.411	18.307
	p-value	0.046	0.040	0.040	0.037	0.037	0.042	0.044	0.048	0.050
C_1	Critical value	8.885	9.200	9.488	9.522	9.409	9.178	9.618	9.351	9.906
	p-value	0.056	0.047	0.041	0.040	0.042	0.048	0.038	0.044	0.050
C_2	Critical value	34.307	34.379	34.146	33.755	33.170	33.543	32.669	32.303	31.410
	p-value	0.024	0.024	0.025	0.028	0.032	0.029	0.037	0.040	0.050

approximate the null distributions fairly well for skewness statistics rather than kurtosis statistics or combined statistics. This can be explained by the slow normal approximation for one dimensional kurtosis in (2.10).

3. Example and Simulation Results

3.1. Example

We consider the data set of Rao (1948) that consists of the thickness of bark deposit on 28 cork trees measured by the weight of cork borings from the north (N), east (E), west (W), and south (S). He

Table 6: Simulated critical values and p-values using asymptotic distribution for $p = 10$, $\alpha = 0.10$

Statistic	Values	$n=30$	$n=40$	$n=50$	$n=100$	$n=150$	$n=200$	$n=250$	$n=300$	$n=\infty$
B_1	Critical value	1.190	0.943	0.780	0.394	0.267	0.201	0.157	0.132	
B_1^*	Critical value	7.247	7.288	7.325	6.973	6.950	6.899	6.709	6.738	6.551
	p-value	0.069	0.067	0.066	0.080	0.081	0.083	0.092	0.090	0.100
S_1	Critical value	6.455	6.620	6.524	6.687	6.732	6.593	6.508	6.538	6.551
	p-value	0.105	0.096	0.101	0.093	0.091	0.098	0.102	0.101	0.100
S_2	Critical value	16.177	6.198	15.814	15.910	16.131	16.008	16.059	16.136	15.987
	p-value	0.095	0.094	0.105	0.102	0.096	0.099	0.098	0.096	0.100
B_2	Critical value	5.613	4.661	3.800	2.035	1.338	0.954	0.782	0.618	
B_2^{**}	Critical value	11.445	11.232	10.642	9.841	9.239	8.568	8.647	8.118	6.551
	p-value	0.007	0.008	0.011	0.017	0.023	0.034	0.032	0.043	0.100
K_1	Critical value	6.665	6.884	6.983	7.168	6.993	7.208	7.026	6.978	6.551
	p-value	0.094	0.084	0.079	0.072	0.079	0.070	0.078	0.080	0.100
K_2	Critical value	15.955	16.316	16.494	16.726	16.623	16.346	16.475	16.067	15.987
	p-value	0.101	0.091	0.086	0.081	0.083	0.090	0.087	0.098	0.100
C_1	Critical value	7.461	7.700	7.974	8.038	8.082	7.928	8.096	7.893	7.790
	p-value	0.119	0.105	0.091	0.088	0.086	0.093	0.085	0.095	0.100
C_2	Critical value	29.932	30.656	30.067	29.796	29.440	29.495	29.267	29.117	28.412
	p-value	0.071	0.060	0.069	0.073	0.079	0.078	0.083	0.085	0.100

Table 7: Statistics and p-values for the contrasts of Rao's bark deposit data

	B_1	S_1	S_2	B_2	K_1	K_2	C_1	C_2
Statistic	0.333	2.004	3.680	0.943	2.378	4.212	2.378	7.892
p-value	0.516	0.495	0.451	0.532	0.409	0.378	0.650	0.444

selected three constraints to investigate if the thickness of bark deposit varies in the four directions,

$$Y_1 = N - E - W + S, \qquad Y_2 = S - W, \qquad Y_3 = N - S$$

by the reason he explained in the paper, and tested $E(Y_i) = 0, i = 1, 2, 3$. The assumption of multi-normality should be valid to test the problem by applying some techniques like Hotelling's T^2 test. Pearson (1956) also examined the data set in Example 2 in his paper.

For the contrasts (Y_1, Y_2, Y_3), the skewness is $(0.662, 0.228, -0.038)$, and the kurtosis is $(3.812, 3.935, 1.641)$. After transforming to $\mathbf{Z}_i$ in (2.1), the skewness in (2.2), and the kurtosis in (2.9) are

$$\left(\sqrt{b_1(1)},\ \sqrt{b_1(2)},\ \sqrt{b_1(3)}\right) = (-0.577, -0.476, -0.21),$$

and

$$(b_2(1), b_2(2), b_2(3)) = (3.764, 2.679, 1.981).$$

Applying the transformations in (2.8) and (2.13) yields

$$\left(T_1^*, T_2^*, T_3^*\right) = (-1.416, -1.180, -0.533), \quad \text{and} \quad \left(T_1^**, T_2^**, T_3^**\right) = (1.353, 0.051, -1.542).$$

Table 7 provides the statistics in this paper and the p-values. The p-values are computed using χ^2 distributions. From the result, the multinormality of the contrasts could not be rejected that confirms Rao's test for contrasts as valid. Mardia (1975), and Srivastava and Hui (1987) also considered the data set and have the same conclusion for the contrasts; however, they have different results for the original data.

Table 8: Power comparison of $b^*_{1,M}, B_1, S_1, S_2, b^{2*}_{2,M}, B_2, K_1, K_2$, and C_1, C_2 ($\alpha = 0.05$, $d = 2$, $n = 20$)

Alternative	Group I				Group II				Group III	
	$b^*_{1,M}$	B_1	S_1	S_2	$b^{2*}_{2,M}$	B_2	K_1	K_2	C_1	C_2
$N(0,1)^2$	5	5	4	5	5	5	4	4	4	4
$C(0,1)^2$	92	91	89	93	**96**	**96**	94	95	95	**96**
Logis$(0,1)^2$	16	12	14	15	15	**18**	10	10	14	15
$(t_2)^2$	64	61	64	63	**71**	69	63	65	68	68
$(t_5)^2$	28	25	24	27	27	**31**	22	22	23	25
$B(1,1)^2$	1	1	0	1	0	0	16	**18**	12	6
$B(2,2)^2$	2	1	2	1	1	1	6	**7**	6	2
$B(1,2)^2$	7	8	8	9	3	2	**10**	8	**10**	6
$\exp(1)^2$	**71**	69	70	**71**	55	48	41	47	64	64
LN$(0,.5)^2$	53	50	51	**57**	41	37	36	32	45	50
$\Gamma(0.5,1)^2$	**91**	87	86	**91**	74	72	62	68	87	85
$\Gamma(5,1)^2$	**24**	22	23	23	17	15	13	14	21	23
$(\chi^2_5)^2$	39	39	38	**43**	29	25	23	20	35	37
$(\chi^2_{15})^2$	17	18	18	**20**	15	13	10	12	14	16
$N(0,1)*t_5$	16	**19**	15	17	18	18	15	15	17	16
$N(0,1)*B(1,1)$	2	3	2	2	3	2	**16**	**16**	12	7
$N(0,1)*\exp(1)$	**47**	43	43	43	31	30	21	23	39	35
$N(0,1)*\chi^2_5$	24	33	**34**	29	19	18	14	16	29	26
NMIX$_2(.5,2,0,0)$	3	3	2	3	3	4	10	**11**	8	4
NMIX$_2(.5,4,0,0)$	3	4	4	4	2	3	**75**	70	68	50
NMIX$_2(.5,2,.9,0)$	**22**	14	14	16	16	21	18	21	**22**	20
NMIX$_2(.5,.5,.9,0)$	14	16	17	18	18	**22**	15	17	19	19
NMIX$_2(.5,.5,.9,-.9)$	37	34	36	36	35	36	27	31	**38**	35

3.2. Simulation results

A simulation was performed to study the power of the proposed tests based on the skewness and kurtosis of the scaled residuals. For dimensions and sample sizes, $d = 2$, $n = 20, 50$, $d = 5$, $n = 20, 50$, and $d = 10$, $n = 50, 100$, the power of the statistics is presented in Tables 8–13 under the significance level $\alpha = 0.05$. We generated 1,000 samples from each of various alternative distributions. The alternatives included in the study are distributions with independent marginals and mixtures of normal distributions. The following notations are used. $N(0,1)$, $C(0,1)$, Logis$(0,1)$, and $\exp(1)$ stand for a standard normal, Cauchy, logistic, and exponential distribution; χ^2_k and t_k are for the chi-square, t-distribution with k degrees of freedom, respectively; $\Gamma(a,b)$ is for the gamma distribution with density $b^{-a}\Gamma(a)^{-1}x^{a-1}\exp(-x/b)$, $x > 0$; $B(a,b)$ is for the beta distribution with density $B(a,b)^{-1}x^{a-1}(1-x)^{b-1}$, $0 < x < 1$; LN(a,b) is for the lognormal distribution with density $(\sqrt{2\pi}bx)^{-1}\exp(-(\log x - a)^2/2b^2)$, $x > 0$. Here $N(0,1)$, $C(0,1)$, Logis$(0,1)$, t_k, $B(1,1)$, and $B(2,2)$ are symmetric distributions, and the others are skewed distributions. The product of k independent copies of F_1 is denoted by F_1^k. The alternatives F_1^k with symmetric marginals F_1 such as $C(0,1)$, Logis$(0,1)$, t_k, $B(1,1)$, and $B(2,2)$ are in the first part of each table. $F_1 * F_2$ denotes the distribution having independent marginal distributions F_1 and F_2. NMIX$_2(\kappa,\delta,\rho_1,\rho_2)$ is for the bivariate normal mixture

$$\kappa N_2\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1 & \rho_1\\ \rho_1 & 1\end{pmatrix}\right) + (1-\kappa)N_2\left(\begin{pmatrix}\delta\\ \delta\end{pmatrix},\begin{pmatrix}1 & \rho_2\\ \rho_2 & 1\end{pmatrix}\right),$$

which is a correlated distribution.

In the study, we put the power of Kim (2004b)'s statistics $b^*_{1,M}$, $b^{2*}_{2,M}$ for comparison. As we mentioned in Section 2, they are approximate statistics to Malkovich and Afifi (1973)'s skewness in

Table 9: Power comparison of $b^*_{1,M}, B_1, S_1, S_2, b^{2*}_{2,M}, B_2, K_1, K_2$, and C_1, C_2 ($\alpha = 0.05$, $d = 2$, $n = 50$)

Alternative	Group I				Group II				Group III	
	$b^*_{1,M}$	B_1	S_1	S_2	$b^{2*}_{2,M}$	B_2	K_1	K_2	C_1	C_2
$N(0,1)^2$	5	4	6	4	4	5	6	6	5	5
$C(0,1)^2$	99	98	98	99	**100**	**100**	**100**	**100**	**100**	**100**
$\text{Logis}(0,1)^2$	22	20	20	20	**33**	26	20	23	22	27
$(t_2)^2$	87	85	86	86	**98**	95	93	95	92	94
$(t_5)^2$	39	38	36	40	**56**	48	41	44	45	49
$B(1,1)^2$	0	0	0	0	0	0	50	**57**	44	37
$B(2,2)^2$	0	0	0	1	0	0	23	**27**	18	10
$B(1,2)^2$	17	18	18	**25**	1	1	18	18	20	20
$\exp(1)^2$	**100**	96	97	97	86	80	72	73	93	96
$\text{LN}(0,.5)^2$	**95**	89	89	90	73	70	60	63	84	84
$\Gamma(0.5,1)^2$	**100**	**100**	99	**100**	96	94	92	95	99	99
$\Gamma(5,1)^2$	**60**	53	54	**60**	29	28	24	23	44	48
$(\chi^2_5)^2$	**85**	81	76	82	53	49	39	43	72	75
$(\chi^2_{15})^2$	43	41	40	**45**	25	22	15	17	33	36
$N(0,1) * t_5$	27	28	22	25	33	**34**	27	28	28	30
$N(0,1) * B(1,1)$	3	2	3	2	2	2	**72**	**72**	66	54
$N(0,1) * \exp(1)$	**93**	82	82	82	62	55	43	46	74	73
$N(0,1) * \chi^2_5$	61	**75**	**75**	73	32	35	30	30	65	59
$\text{NMIX}_2(.5,2,0,0)$	2	3	3	2	3	3	**39**	38	31	25
$\text{NMIX}_2(.5,4,0,0)$	2	2	2	3	2	28	**99**	**99**	**99**	**99**
$\text{NMIX}_2(.5,2,.9,0)$	51	19	19	22	33	45	44	**52**	40	42
$\text{NMIX}_2(.5,.5,.9,0)$	26	24	23	26	35	**41**	28	33	32	34
$\text{NMIX}_2(.5,.5,.9,-.9)$	53	53	51	54	**66**	60	45	56	54	64

(2.5) and kurtosis in (2.12), respectively. The statistics $b^*_{1,M}$, $b^{2*}_{2,M}$ are as follows,

$$b^*_{1,M} = \max_{1\le l\le n} \frac{1}{n^2} \frac{\left[\sum_{j=1}^n \left(\left(\boldsymbol{X}_l - \bar{\boldsymbol{X}}\right)' \boldsymbol{S}^{-1} \left(\boldsymbol{X}_j - \bar{\boldsymbol{X}}\right)\right)^3\right]^2}{\left[\left(\boldsymbol{X}_l - \bar{\boldsymbol{X}}\right)' \boldsymbol{S}^{-1} \left(\boldsymbol{X}_l - \bar{\boldsymbol{X}}\right)\right]^3},$$

$$b^{2*}_{2,M} = \max_{1\le l\le n} \left[\frac{1}{n} \frac{\sum_{j=1}^n \left(\left(\boldsymbol{X}_l - \bar{\boldsymbol{X}}\right)' \boldsymbol{S}^{-1} \left(\boldsymbol{X}_j - \bar{\boldsymbol{X}}\right)\right)^4}{\left[\left(\boldsymbol{X}_l - \bar{\boldsymbol{X}}\right)' \boldsymbol{S}^{-1} \left(\boldsymbol{X}_l - \bar{\boldsymbol{X}}\right)\right]^2} - 3\right]^2.$$

From now on, we refer to Group I for statistics based on skewness, $b^*_{1,M}$, B_1, S_1, and S_2, Group II for $b^*_{2,M}$, B_2, K_1, and K_2 on kurtosis, and Group III for C_1 and C_2 that are the combined statistics shown in Tables 8–13.

Each number in Tables 8–13 represents the empirical power of each test in percentage form rounded to the next integer. From the first line of each table, we can see that every statistic seems to have good control of type I error due to the use of critical values from the simulation rather than from asymptotic distributions. The best power for each alternative is written in bold, except when the numbers are all the same as 100.

The power study in Tables 8–13 indicates the following. First, statistics in each group show similar power, except that K_1, K_2 are more powerful when alternatives are $\text{NMIX}_2(.5,2,0,0)$, $\text{NMIX}_2(.5,4,0,0)$, or with shorter tailed marginals beta distributions such as $B(1,1)^k$, $B(2,2)^k$, $B(1,2)^k$, and $N(0,1)^{k-1} * B(1,1)$, $k = 2,5,10$. This happens for every combination of d and n considered in the study, and appears predominantly when the sample size n is big relative to the dimension p. Statistics that belong

Table 10: Power comparison of $b^{*}_{1,M}, B_1, S_1, S_2, b^{2*}_{2,M}, B_2, K_1, K_2$, and C_1, C_2 ($\alpha = 0.05$, $d = 5$, $n = 20$)

Alternative	Group I				Group II				Group III	
	$b^{*}_{1,M}$	B_1	S_1	S_2	$b^{2*}_{2,M}$	B_2	K_1	K_2	C_1	C_2
$N(0,1)^5$	4	6	5	4	6	6	6	6	4	4
$C(0,1)^5$	98	97	98	99	**100**	**100**	99	**100**	99	99
Logis$(0,1)^5$	12	12	13	14	**16**	14	12	11	11	14
$(t_2)^5$	78	75	76	81	81	78	76	77	78	**85**
$(t_5)^5$	28	26	24	25	**30**	28	23	25	26	28
$B(1,1)^5$	1	2	2	2	1	3	6	**9**	6	3
$B(2,2)^5$	1	2	3	3	2	2	**6**	**6**	4	2
$B(1,2)^5$	3	5	5	**7**	4	4	**7**	**7**	6	6
$\exp(1)^5$	59	61	59	**69**	57	49	42	45	55	64
LN$(0,.5)^5$	44	44	48	**56**	44	42	36	38	44	49
$\Gamma(0.5,1)^5$	83	82	82	**89**	81	74	62	74	80	88
$\Gamma(5,1)^5$	15	18	20	**21**	19	15	11	13	17	16
$(\chi^2_5)^5$	30	32	31	**35**	28	27	23	20	29	**35**
$(\chi^2_{15})^5$	11	13	13	**18**	14	12	8	10	12	14
$N(0,1)^4 * t_5$	9	11	11	11	10	**12**	10	11	11	**12**
$N(0,1)^4 * B(1,1)$	4	4	3	4	4	4	5	**8**	5	4
$N(0,1)^4 * \exp(1)$	**21**	18	17	18	18	14	14	12	16	15
$N(0,1)^4 * \chi^2_5$	11	**24**	22	22	11	17	11	14	19	17

Table 11: Power comparison of $b^{*}_{1,M}, B_1, S_1, S_2, b^{2*}_{2,M}, B_2, K_1, K_2$, and C_1, C_2 ($\alpha = 0.05$, $d = 5$, $n = 50$)

Alternative	Group I				Group II				Group III	
	$b^{*}_{1,M}$	B_1	S_1	S_2	$b^{2*}_{2,M}$	B_2	K_1	K_2	C_1	C_2
$N(0,1)^5$	6	4	4	4	5	5	5	4	5	6
$C(0,1)^5$	100	100	100	100	100	100	100	100	100	100
Logis$(0,1)^5$	**30**	18	18	20	**30**	26	17	18	18	24
$(t_2)^5$	98	95	94	96	**100**	99	98	**100**	98	99
$(t_5)^5$	51	40	42	46	**60**	55	43	46	43	54
$B(1,1)^5$	0	2	1	1	0	1	22	**25**	16	10
$B(2,2)^5$	1	1	1	1	0	1	11	**12**	8	4
$B(1,2)^5$	2	7	7	**11**	1	2	9	10	8	8
$\exp(1)^5$	**99**	94	91	97	91	82	73	81	91	96
LN$(0,.5)^5$	**94**	87	86	91	82	74	64	70	79	89
$\Gamma(0.5,1)^5$	**100**	99	**100**	**100**	99	98	94	98	99	**100**
$\Gamma(5,1)^5$	**50**	41	39	49	34	28	22	20	36	44
$(\chi^2_5)^5$	**80**	67	67	75	57	46	36	40	64	69
$(\chi^2_{15})^5$	**39**	29	30	36	27	17	15	16	24	32
$N(0,1)^4 * t_5$	19	17	16	18	22	**25**	20	21	22	21
$N(0,1)^4 * B(1,1)$	4	3	4	4	3	4	**42**	37	34	24
$N(0,1)^4 * \exp(1)$	**59**	39	40	48	40	27	26	21	36	37
$N(0,1)^4 * \chi^2_5$	32	61	**62**	58	19	26	25	22	55	47

to Group I or Group II show very poor power against the above alternatives; however, K_1, K_2 in Group II and C_1, C_2 in Group III have relatively good power, and K_1, K_2 is slightly superior to C_1, C_2 against the alternatives.

Second, Group I statistics based on skewness seem to have better power than Group II statistics based on kurtosis against alternatives with skewed marginal distributions such as $\exp(1)^k$, LN$(0,.5)^k$, $\Gamma(.5,1)^k$, $\Gamma(5,1)^k$, $(\chi^2_5)^k$, $(\chi^2_{15})^k$, $N(0,1)^{k-1} * \exp(1)$, and $N(0,1)^{k-1} * \chi^2_5$. Group I have similar or better power to Group III combined statistics against these alternatives.

Table 12: Power comparison of $B_1, S_1, S_2, B_2, K_1, K_2$, and C_1, C_2 ($\alpha = 0.05, d = 10, n = 50$)

Alternative	Group I			Group II			Group III	
	B_1	S_1	S_2	B_2	K_1	K_2	C_1	C_2
$N(0,1)^{10}$	4	6	5	4	4	4	5	5
$C(0,1)^{10}$	100	100	100	100	100	100	100	100
$\text{Logis}(0,1)^{10}$	16	**18**	17	16	12	15	12	**18**
$(t_2)^{10}$	98	98	**100**	99	99	**100**	99	**100**
$(t_5)^{10}$	44	44	48	**51**	40	45	43	50
$B(1,1)^{10}$	2	2	2	2	7	**10**	6	4
$B(2,2)^{10}$	2	2	2	2	**7**	6	6	3
$B(1,2)^{10}$	4	5	**7**	3	5	4	**7**	5
$\exp(1)^{10}$	87	88	**95**	76	61	77	81	92
$\text{LN}(0,.5)^{10}$	81	80	**87**	69	58	68	73	85
$\Gamma(0.5,1)^{10}$	98	98	**100**	96	90	97	97	**100**
$\Gamma(5,1)^{10}$	29	31	**41**	21	12	17	24	28
$(\chi^2_5)^{10}$	54	54	**66**	38	27	33	43	57
$(\chi^2_{15})^{10}$	23	24	**28**	15	12	11	15	20
$N(0,1)^9 * t_5$	14	15	15	**17**	14	14	14	15
$N(0,1)^9 * B(1,1)$	4	4	5	4	**13**	12	11	8
$N(0,1)^9 * \exp(1)$	22	**23**	19	16	12	12	16	18
$N(0,1)^9 * \chi^2_5$	53	**54**	45	23	18	17	41	34

Table 13: Power comparison of $B_1, S_1, S_2, B_2, K_1, K_2$, and C_1, C_2 ($\alpha = 0.05, d = 10, n = 100$)

Alternative	Group I			Group II			Group III	
	B_1	S_1	S_2	B_2	K_1	K_2	C_1	C_2
$N(0,1)^{10}$	5	4	5	4	5	5	6	5
$C(0,1)^{10}$	100	100	100	100	100	100	100	100
$\text{Logis}(0,1)^{10}$	20	16	25	**29**	15	22	18	27
$(t_2)^{10}$	100	100	100	100	100	100	100	100
$(t_5)^{10}$	57	56	62	75	64	72	63	**76**
$B(1,1)^{10}$	1	1	1	1	19	**21**	14	7
$B(2,2)^{10}$	2	2	1	2	9	**12**	8	4
$B(1,2)^{10}$	8	7	**12**	2	8	8	9	10
$\exp(1)^{10}$	99	97	**100**	95	88	95	98	**100**
$\text{LN}(0,.5)^{10}$	96	96	**99**	88	80	91	94	98
$\Gamma(0.5,1)^{10}$	100	100	100	100	100	100	100	100
$\Gamma(5,1)^{10}$	52	51	**67**	32	20	23	41	56
$(\chi^2_5)^{10}$	81	79	**90**	58	43	55	73	88
$(\chi^2_{15})^{10}$	33	33	**49**	24	15	16	27	38
$N(0,1)^9 * t_5$	21	22	20	**29**	27	23	26	25
$N(0,1)^9 * B(1,1)$	6	4	3	4	**67**	56	65	41
$N(0,1)^9 * \exp(1)$	33	31	**34**	21	18	17	28	30
$N(0,1)^9 * \chi^2_5$	**93**	92	80	43	31	28	86	64

Third, we see that the statistic $b^{2*}_{2,M}$ is more sensitive than the other statistics in Group II, B_2, K_1, K_2, except the alternatives with beta marginals or some normal mixtures. This might be explained by the fact that $b^{2*}_{2,M}$ is trying to search the direction of vector that gives at least normal to the univariate one. For alternatives with beta marginals, K_1 and K_2 show superior power.

In general, it seems Group III statistics C_1, C_2 have relatively good power for all alternatives under study. They are likely to have the merits of both Group I and Group II, and they might be considered omnibus tests, although they are not uniformly the most powerful, and the sample size needs to be moderately large to use asymptotic distribution.

4. Concluding Remarks

In this study, we propose test statistics based on the skewness and kurtosis of the scaled residuals for testing multinormality. We consider the skewness or the kurtosis in each coordinate of the scaled residuals, and transform them to normality in each coordinate. The null distribution of the statistic can be approximated by simple distribution, and the approximation is adequate by applying the transformation to normality. Through the simulation study, the combined statistics of skewness and kurtosis show moderate sensitivity for all alternatives under study.

References

Anscombe, F. J. and Glynn, W. J. (1983). Distribution of the kurtosis statistic b_2 for normal samples, *Biometrika*, **70**, 227–234.

Baringhaus, L. and Henze, N. (1992). Limit distributions for Mardia's measure of multivariate skewness, *Annals of Statistics*, **20**, 1889–1902.

Bowman, K. O. and Shenton, L. R. (1975). Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2, *Biometrika*, **62**, 243–250.

Bowman, K. O. and Shenton, L. R. (1986). Moment ($\sqrt{b_1}, b_2$) Techniques, In R. B. D'Agostino and M. A. Stephens (Eds.), Goodness-of-Fit Techniques, Marcel Dekker, New York, 279–329.

D'Agostino, R. B. (1970). Transformation to normality of the null distribution of g_1, *Biometrika*, **57**, 679–681.

D'Agostino, R. B. (1986). Tests for the normal distribution. In R. B. D'Agostino and M. A. Stephens (Eds.), Goodness-of-Fit Techniques, Marcel Dekker, New York, 367–420.

D'Agostino, R. B. and Pearson, E. S. (1973). Tests for departure from normality: Empirical results for the distributions of b_2 and $\sqrt{b_1}$, *Biometrika*, **60**, 613–622.

D'Agostino, R. B. and Pearson, E. S. (1974). Correction and amendment: Tests for departure from normality: Empirical results for the distributions of b_2 and $\sqrt{b_1}$, *Biometrika*, **61**, 647.

De Wet, T. and Venter, J. H. (1972). Asymptotic distributions of certain test criteria of normality, *South African Statistical Journal*, **6**, 135–149.

Farrell, P. J., Salibian-Barrera, M. and Naczk, K. (2007). On tests for multivariate normality and associated simulation studies, *Journal of Statistical Computation and Simulation*, **77**, 1065–1080.

Fattorini, L. (1986). Remarks on the use of the Shapiro-Wilk statistic for testing multivariate normality, *Statistica*, **46**, 209–217.

Henze, N. (1994). On Mardia's kurtosis for multivariate normality, *Communications in Statistics-Theory and Methods*, **23**, 1031–1045.

Henze, N. (2002). Invariant tests for multivariate normality: A critical review, *Statistical Papers*, **43**, 467–506.

Henze, N. and Zirkler, B. (1990). A class of invariant consistent tests for multivariate normality, *Communications in Statistics-Theory and Methods*, **19**, 3539–3617.

Horswell, R. L. and Looney, S. W. (1992). A comparison of tests for multivariate normality that are based on measures of multivariate skewness and kurtosis, *Journal of Statistical Computation and Simulation*, **42**, 21–38.

Kendall, M. and Stuart, A. (1977). *The Advanced Theory of Statistics, Vol. I*, MacMillan Publishing Co., New York.

Kim, N. (2004a). An approximate Shapiro-Wilk statistic for testing multivariate normality, *The Korean Journal of Applied Statistics*, **17**, 35–47.

Kim, N. (2004b). Remarks on the use of multivariate skewness and kurtosis for testing multivariate normality, *The Korean Journal of Applied Statistics*, **17**, 507–518.

Kim, N. (2005). The limit distribution of an invariant test statistic for multivariate normality, *The Korean Communications in Statistics*, **12**, 71–86.

Kim, N. and Bickel, P. J. (2003). The limit distribution of a test statistic for bivariate normality, *Statistica Sinica*, **13**, 327–349.

Malkovich, J. F. and Afifi, A. A. (1973). On tests for multivariate normality, *Journal of the American Statistical Association*, **68**, 176–179.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications, *Biometrika*, **57**, 519–530.

Mardia, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis for testing normality and robustness studies, *Sankhyā: The Indian Journal of Statistics Series B*, **36**, 115–128.

Mardia, K. V. (1975). Assessment of multinormality and the robustness of Hotelling's T^2 test, *Applied Statistics*, **24**, 163–171.

Mecklin, C. J. and Mundfrom, D. J. (2005). A Monte Carlo comparison of the Type I and Type II error rates of tests of multivariate normality, *Journal of Statistical Computation and Simulation*, **75**, 93–107.

Mudholkar, G. S., Srivastava, D. K. and Lin, C. T. (1995). Some p-variate adaptations of the Shapiro-Wilk test of normality, *Communications of Statistics-Theory and Methods*, **24**, 953–985.

Pearson, E. S. (1956). Some aspects of the geometry of statistics, *Journal of the Royal Statistical Society Series A (General)*, **119**, 125–146.

Pearson, E. S., D'Agostino, R. B. and Bowman, K. O. (1977). Tests for departure from normality: Comparison of powers, *Biometrika*, **64**, 231–246.

Rao, C. R. (1948). Test of significance in multivariate analysis, *Biometrika*, **35**, 58–79.

Romeu, J. L. and Ozturk, A. (1993). A comparative study of goodness-of-fit tests for multivariate normality, *Journal of Multivariate Analysis*, **46**, 309–334.

Roy, S. N. (1953). On a heuristic method of test construction and its use in multivariate analysis, *Annals of Mathematical Statistics*, **24**, 220–238.

Royston, J. P. (1983). Some techniques for accessing multivariate normality based on the Shapiro-Wilk W, *Applied Statistics*, **32**, 121–133.

Shapiro, S. S. and Francia, R. S. (1972). An approximate analysis of variance test for normality, *Journal of the American Statistical Association*, **67**, 215–216.

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples), *Biometrika*, **52**, 591–611.

Srivastava, D. K. and Mudholkar, G. S. (2003). Goodness of fit tests for univariate and multivariate normal models, *Handbook of Statistics*, **22**, 869–906.

Srivastava, M. S. (1984). A measure of skewness and kurtosis and a graphical method for assessing multivariate normality, *Statistics & Probability Letters*, **2**, 263–267.

Srivastava, M. S. and Hui, T. K. (1987). On assessing multivariate normality based on Shapiro-Wilk *W* statistic, *Statistics & Probability Letters*, **5**, 15–18.

Thode, H. C. (2002). *Testing for Normality*, Marcel Dekker, New York.

Villasenor Alva, J. A. and González Estrada, E. (2009). A generalization of Shapiro-Wilk's test for multivariate normality, *Communications in Statistics-Theory and Methods*, **38**, 1870–1833.

Received April 25, 2015; Revised June 9, 2015; Accepted June 9, 2015