Proposal and Pilot Study: A Generalization of the W or W' Statistic for Multivariate Normality

José Moral-De La Rubia

School of Psychology, Universidad Autónoma de Nuevo León, Nuevo León, Monterrey, México
Email: jose.morald@uanl.edu.mx, jose_moral@hotmail.com

Abstract

The aim of this paper is to present a generalization of the Shapiro-Wilk W-test or Shapiro-Francia W'-test for application to two or more variables. It consists of calculating all the unweighted linear combinations of the variables and their W- or W'-statistics with the Royston's log-transformation and standardization, $z_{\ln(1-W)}$ or $z_{\ln(1-W')}$. Because the calculation of the probability of $z_{\ln(1-W)}$ or $z_{\ln(1-W')}$ is to the right tail, negative values are truncated to 0 before doing their sum of squares. Independence in the sequence of these half-normally distributed values is required for the test statistic to follow a chi-square distribution. This assumption is checked using the robust Ljung-Box test. One degree of freedom is lost for each cancelled value. Defined the new test with its two variants (Q-test or Q'-test), 50 random samples with 4 variables and 20 participants were generated, 20% following a multivariate normal distribution and 80% deviating from this distribution. The new test was compared with Mardia's, runs, and Royston's tests. Central tendency differences in type II error and statistical power were tested using the Friedman's test and pairwise comparisons using the Wilcoxon's test. Differences in the frequency of successes in statistical decision making were compared using the Cochran's Q test and pairwise comparisons using the McNemar's test. Sensitivity, specificity and efficiency proportions were compared using the McNemar's Z test. The generated 50 samples were classified into five ordered categories of deviation from multivariate normality, the correlation between this variable and p-value of each test was calculated using the Spearman's coefficient and these correlations were compared. Family-wise error rate corrections were applied. The new test and the Royston's test were the best choices, with a very slight advantage Q-test over Q'-test. Based on these

promising results, further study and use of this new sensitive, specific and effective test are suggested.

Keywords

1. Introduction

The present article aims to: 1) to present a new multivariate normality test as a generalization through a sum of squares of the Shapiro-Wilk W statistic [1] or the Shapiro-Francia W' statistic [2] with the Royston's log-transformations and standardizations [3] [4]; as well as 2) to compare the central tendency of type II error or statistical power; 3) the frequency of successes and errors in making the decision on the null hypothesis, 4) the sensitivity (proportion of detection of multivariate normality cases), specificity (proportion of rejection of non-cases) and efficiency (proportion of correctness in classifying), and 5) the correlation between the critical level or probability value and the deviation from multivariate normality (variable of five ordered categories defined when classifying the generated samples). The two versions of the new test are compared with the Mardia's K^2-test based on multivariate skewness and kurtosis [5] [6] [7], the Friedman-Rafsky multivariate runs test [8] applied to the multivariate normality test by Smith and Jain [9] and the Royston's H-test [10], either from the W-statistics [3] or W'-statistics [4]. It is a very simple test both in its rationale and application, which is very easy to learn and teach. It can be run with the Excel program and could be included in the multivariate normality library of the R program, making it accessible to undergraduate students of social sciences, biosciences, and other empirical sciences with a stronger mathematical emphasis.

Why a new statistical test when there are already several for this purpose? Because it constitutes a good option despite the simplicity of its rationale, being a Q-type test of sum of variables with standard half-normal distribution (standard normal truncated from quantile 0.5 to quantile 1) that assumes independence in the sequence of summed values to use the approximation to the chi-squared distribution in the calculation of the critical level or probability value [11] [12] [13], and checks this assumption of independency in a specified sequence that allows estimating autocorrelation in case of non-compliance for use in the calculation of bootstrap-based critical values.

2. The Proposed Q Test from the W Statistics or Q' Test from the W' Statistics

The test is based on the lemma or proven proposition that, if a set of k variables comes from a multivariate normal distribution, any linear combination among

the variables follows a univariate normal distribution [14]. The lemma that the sum of squares of k independent variables with standard normal distribution follows a chi-square distribution with k degrees of freedom was initially considered [15]. In addition, a random sample of size k is defined as a succession of k random, independent and identically distributed variables [11].

2.1. Formulation of Statistical Hypotheses

Let there be a random sample of n participants or elements to whom k measurements have been made with the same variable X or with k different variables. It is assumed that the variables are continuous quantitative or admit a continuous generalization so that they can fit a normal distribution model.

The null hypothesis (H_0) states that the random vector $\vec{x}$ composed of k random variables follows a multivariate normal distribution of unknown parameters: $\vec{\mu}$ (vector of population means) and Σ (population covariance matrix).

$$H_0 : \vec{x} \sim N\left(\vec{\mu}, \Sigma\right)$$

$$\vec{x}, \vec{\mu} \in \mathbb{R}^k \quad \text{y} \quad \Sigma \in \mathbb{R}^{k \times k}$$

The alternative hypothesis (H_1) posits that the random vector $\vec{x}$ does not follow a multivariate normal distribution of unknown parameters $\vec{\mu}$ and Σ.

$$H_1 : \vec{x} \nsim N\left(\vec{\mu}, \Sigma\right)$$

2.2. Assumptions

- n independent samples of k tuples, that is, a random sample of n participants with scores on k continuous quantitative variables that may be correlated or independent.
- Large sample size n, at least 20 participants.
- Serial independence between the $2^k - 1$ values of the log-transformed, standardized and truncated W or W' statistics, considered as a succession of identically distributed random variables. The sequence is specified by the subscript of the variable (X_1, X_2, …, X_k) and the increasing complexity of the unweighted linear combination (one variable, two variables, … k variables).

2.3. Test Statistic

To obtain the test statistic Q (from the Shapiro-Wilk W statistics) or Q' (from the Shapiro-Francia W' statistics), the following steps are followed:

1) All unweighted linear combinations (by simple summation) among the k variables are obtained and the variables of each combination are summed, which results in k combinations of one variable, $k(k-1)/2$ sums of combinations of two variables, $_kC_3$ (non-repeating combination of k variables taken in groups of three) sums of combinations of three variables, …, and a sum of the k variables, that is, $2^k - 1$ sums of linear combinations of k variables.

$$\sum_{j=1}^{k} \binom{k}{j} = \binom{k}{1} + \binom{k}{2} + \cdots + \binom{k}{k} = 2^k - 1$$

2) A first alternative is to calculate the Q-statistic that starts from the standardized values of the log-transformed W-statistics [3] of each of the $2^k - 1$ sums of linear combinations of the k variables. The Shapiro-Wilk W statistic is the square of the correlation between the empirical quantiles $x_{(i)}$ and the standardized and normalized theoretical quantiles a_i.

$$W_l = r^2_{x_{(i)l},a} = \frac{\left[\sum_{i=1}^{n}\left(x_{(i)l} - \overline{x}_l\right)\left(a_i - \overline{a}\right)\right]^2}{\sum_{i=1}^{n}\left(x_{(i)l} - \overline{x}_l\right)^2 \sum_{i=1}^{n}\left(a_i - \overline{a}\right)^2} = \frac{\left[\sum_{i=1}^{n}\left(x_{(i)l} - \overline{x}_l\right)a_i\right]^2}{\sum_{i=1}^{n}\left(x_{(i)l} - \overline{x}_l\right)^2}$$

$$i = 1, 2, \cdots, n \quad \text{y} \quad l = 1, 2, \cdots, 2^k - 1$$

$x_{(i)l}$ = empirical quantiles or scores of the variable X_l (l = 1, 2, ..., 2^{k-1}) sorted in ascending order, from 1 to n.

$\text{x}_{(i)l}$: $x_{(1)l}$ $x_{(2)l}$ ... $x_{(n)l}$

(i): 1 2 ... n

$\overline{x}_l = \sum_{i=1}^{n} x_{li}/n$ = the arithmetic mean of variable X_l.

The n expected values under a standard normal distribution, denoted by m_i, are common for the 2^{k-1} sums of the linear combinations of the k variables, as are the n standardized and normalized values, denoted by a_i. To obtain the a_i, the m_i must first be calculated.

$$m_i = \Phi^{-1}\left(\frac{(i) - a}{n + 1 - a - b}\right) = \Phi^{-1}\left(\frac{(i) - 0.375}{n + 0.25}\right)$$

A value of 3/8 is given to the a and b values. It is based on the fact that the i order statistic of a variable with a standard uniform distribution follows a beta distribution of parameters: $\alpha = i$ and $\beta = n + 1 - i$, whose expected value is $\alpha/(\alpha + \beta) = i/(n + 1)$. The values a and b can vary from 0 to 1 and 3/8 achieves the estimates with the best approximation to the quantiles of a normal distribution for many different sample sizes [16].

$$m = \sum_{i=1}^{n} m_i^2$$

$$u = 1/\sqrt{n}$$

$$a_n = m_n/\sqrt{m} + 0.221157u - 0.147981u^2 - 2.071190u^3$$
$$+ 4.434685u^4 - 2.706056u^5$$

$$a_{n-1} = m_{n-1}/\sqrt{m} + 0.042981u - 0.293762u^2 - 1.752461u^3$$
$$+ 5.682633u^4 - 3.582663u^5$$

$$a_1 = -a_n$$

$$a_2 = -a_{n-1}$$

$$\epsilon = \frac{m - 2m_n^2 - 2m_{n-1}^2}{1 - 2a_n^2 - 2a_{n-1}^2}$$

$$a_i = m_i/\sqrt{\epsilon}\,; i = 3, 4, \cdots, n - 2$$

$$\sum_{i=1}^{n} a_i = 0; \quad \bar{a} = \sum_{i=1}^{n} a_i/n = 0; \quad \sum_{i=1}^{n} a_i^2 = 1; \quad \sum_{i=1}^{n} (a_i - \bar{a})^2 = 1$$

The standardized value of the logarithmically transformed statistic for sample sizes from 12 to 5000 ($n \geq 20$) is obtained using the following formulas:

$$Z_{\ln(1-W_l)} = \frac{\ln(1-W_l) - \mu_{\ln(1-W_l)}}{\sigma_{\ln(1-W_l)}} \sim N(0,1)$$

$$\mu_{\ln(1-W_l)} = -1.5861 - 0.31082 \ln(n) - 0.083751 \big[\ln(n)\big]^2 + 0.0038915 \big[\ln(n)\big]^3$$

$$\sigma_{\ln(1-W_l)} = e^{-0.4803 - 0.082676 \ln(n) + 0.0030302 \big[\ln(n)\big]^2}$$

If $P\big(Z \geq z_{\ln(1-W_l)}\big) = 1 - P\big(Z < z_{\ln(1-W_l)}\big) \geq \alpha$, H_0 is accepted

If $P\big(Z \geq z_{\ln(1-W_l)}\big) = 1 - P\big(Z < z_{\ln(1-W_l)}\big) < \alpha$, H_0 is rejected.

The other alternative is to calculate the Q' statistic using the standardized values of the logarithmically transformed Shapiro-Francia W' statistics [4] of the $2^k - 1$ sums of linear combinations among the k variables. The test statistic for univariate normality from Shapiro and Francia [2] is used, which is the square of the correlation between the empirical quantiles and the expected quantiles under normal distribution [4].

$$W_l' = r_{x_{(i)l}m}^2 = \frac{\left[\sum_{i=1}^{n} \big(x_{(i)l} - \bar{x}_l\big)\big(m_i - \bar{m}\big)\right]^2}{\sum_{i=1}^{n} \big(x_{il} - \bar{x}_l\big)^2 \sum_{i=1}^{n} \big(a_i - \bar{a}\big)^2} = \frac{\left[\sum_{i=1}^{n} \big(x_{(i)l} - \bar{x}_l\big)m_i\right]^2}{\sum_{i=1}^{n} \big(x_{il} - \bar{x}_l\big)^2 \sum_{i=1}^{n} m_i^2}$$

m_i = theoretical quantiles that are the same for the k variables.

$$m_i = \Phi^{-1}\left(\frac{(i) - 0.375}{n + 0.25}\right)$$

$$\sum_{i=1}^{n} m_i = 0; \quad \bar{m} = \sum_{i=1}^{n} m_i/n = 0$$

The standardization of the logarithmically transformed W' statistic requires calculating its mean and standard deviation that depend on the u and v values, respectively. The u and v values only depend on the sample size n, so they are common to the $2^k - 1$ transformed statistics, as are the means and standard deviations. This standardization was developed for sample sizes from 5 to 5000.

$$\ln(1 - W_l')$$

$$u = \ln\big[\ln(n)\big] - \ln(n)$$

$$\mu_{\ln(1-W_l')} = 1.0528u - 1.2725$$

$$v = \ln\big[\ln(n)\big] + 2/\ln(n)$$

$$\sigma_{\ln(1-W_l')} = -0.26758v + 1.0308$$

The standardized value of the logarithmic transformation of W' statistic fol-

lows a standard normal distribution [4].

$$Z_{\ln(1-W_l')} = \frac{\ln(1-W_l') - \mu_{\ln(1-W_l')}}{\sigma_{\ln(1-W_l')}} \sim N(0,1)$$

If $P\left(Z \geq z_{\ln(1-W_l')}\right) = 1 - P\left(Z < z_{\ln(1-W_l')}\right) \geq \alpha$, H_0: $X_l \sim N\left(\mu_{X_l}, \sigma_{X_l}^2\right)$ is accepted.

If $P\left(Z \geq z_{\ln(1-W_l')}\right) = 1 - P\left(Z < z_{\ln(1-W_l')}\right) < \alpha$, rechaza H_0 is rejected.

3) Finally, the sum of the squares of the $2^k - 1$ z-statistics is calculated, thus obtaining the value of the test statistic Q or Q'. Since the calculation of the probability is one-sided towards the right tail, a negative value implies a good fit to normality. The more negative it is, the better the fit. Therefore, it is necessary to truncate any negative value; otherwise, the value of Q or Q' would be wrongly inflated. The number of canceled values is denoted by *a*.

$$z_l = \begin{cases} 0 & z_{\ln(1-W_l)}^2 < 0 \\ z_{\ln(1-W_l)}^2 & z_{\ln(1-W_l)}^2 \geq 0 \end{cases} \qquad z_l' = \begin{cases} 0 & z_{\ln(1-W_l')}^2 < 0 \\ z_{\ln(1-W_l')}^2 & z_{\ln(1-W_l')}^2 \geq 0 \end{cases}$$

$$Q = \sum_{l=1}^{2^k-1} z_l^2 \qquad\qquad Q' = \sum_{l=1}^{2^k-1} \left(z_l'\right)^2$$

2.4. Sampling Distribution

Initially, there were $2^k - 1$ identically distributed (with standard normal distribution) random variables with samples of size 1 (inferential reconceptualization of sample values: $z_{\ln(1-W)}$ or $z_{\ln(1-W')}$, whose sum of squares in case of independence follows a chi-square distribution with $2^k - 1$ degrees of freedom. However, once the variables are truncated, they become standard half-normal distribution with mean or mathematical expectation: $E(Z_l) = \sigma \times \sqrt{2/\pi} = \sqrt{2/\pi}$ and variance: $var(Z_l) = \sigma^2 \times (1 - 2/\pi) = (1 - 2/\pi)$, where $\sigma^2 = 1$ [17]. To solve this obstacle, additional lemmas were considered.

If *X* follows a half-normal distribution, the ratio between the square of the variable and σ^2 follows a chi-square distribution with one degree of freedom: $X^2/\sigma^2 \sim \chi_{(1)}^2$. In turn, the sum of $2^k - 1$ independent variables with chi-squared distribution with one degree of freedom follows a chi-squared distribution with $2^k - 1$ degrees of freedom [18], leading to the initial path without the need to make any changes, since the truncated $z_{\ln(1-W)}$ or $z_{\ln(1-W')}$ values are divided by $\sigma^2 = 1$ [17]. It is worth mentioning that the correspondence of the quantiles of a chi-squared distribution with a $\chi_{(1)}^2$ degree of freedom is more direct with the squared standard half-normal distribution (HZ^2) than with the squared standard normal standard distribution (Z^2). This is due to the 0 to $+\infty$ support, positive skewness and leptokurtosis of the chi-square and half-normal distributions versus the $-\infty$ to $+\infty$ support, symmetry and meso-kurtosis of the normal distribution (Table 1).

$$_p\chi_1^2 \equiv HZ_p^2 \qquad _p\chi_1^2 \equiv Z_{p'=0.5+p/2}^2$$

Table 1. Correspondence among the quantiles of a chi-square distribution with one degree of freedom, a squared standard half-normal distribution, and squared standard normal distribution.

p	$_p\chi^2_{(1)}$	HZ_p	HZ^2_p	$p' = 0.5 + p/2$	$Z_{p'}$	$Z^2_{p'}$
0.05	0.004	0.063	0.004	0.525	0.063	0.004
0.1	0.016	0.126	0.016	0.55	0.126	0.016
0.2	0.064	0.253	0.064	0.6	0.253	0.064
0.25	0.102	0.318	0.101	0.625	0.319	0.102
0.3	0.148	0.385	0.148	0.65	0.385	0.148
0.4	0.275	0.524	0.275	0.7	0.524	0.275
0.5	0.455	0.674	0.455	0.75	0.455	0.455
0.6	0.708	0.841	0.708	0.8	0.842	0.708
0.7	1.074	1.036	1.074	0.85	1.036	1.074
0.75	1.323	1.150	1.323	0.875	1.150	1.323
0.8	1.642	1.281	1.642	0.9	1.282	1.642
0.9	2.706	1.645	2.706	0.95	1.645	2.706
0.95	3.841	1.960	3.841	0.975	1.960	3.841
μ	1	0.7979	1		0	1
Var.	2	0.3634	2		1	2
$\sqrt{\beta_1}$	$\sqrt{8}$	0.995272			0	
$\beta_2 - 3$	12	0.869177			0	

Note: p = quantile order, $_p\chi^2_{(1)}$ = the pth quantile of a chi-square distribution with one degree of freedom, HZ_p = the pth quantile of a standard half-normal distribution, HZ^2_p = the pth quantile of a standard half-normal distribution, $p' = 0.5 + p/2$ = modified quantile order, $Z_{p'}$ = the quantile of order p' of a standard normal distribution, $Z^2_{p'}$ = the quantile of order p' of a squared standard normal distribution. Population or distribution statistics: μ = mean or mathematical expectation, Var. = variance, $\sqrt{\beta_1}$ = the measure of skewness based on the third standardized central moment, $\beta_2 - 3$ = excess kurtosis or measure of kurtosis based on the fourth standardized central moment minus the expected value for a normal distribution which is 3.

The test statistic Q or Q' is the sum of squares of $2^k - 1$ random variables with a chi-square distribution with one degree of freedom. In case the $2^k - 1$ random variables are independent, the sampling distribution of Q or Q' statistics would be a chi-square with $2^k - 1$ degrees of freedom [15] [18] [19]. However, a correction for the degrees of freedom was introduced to achieve greater specificity in the test by eliminating one degree of freedom for each canceled variable. Consequently, the degrees of freedom become $2^k - 1 - a$, where a is the number of nulled variables.

$$Q = \sum_{i=1}^{2^k-1} z_l \quad \text{ó} \quad Q' = \sum_{i=1}^{2^k-1} z_l' \sim \chi^2_{2^k-1-a}$$

In case the independence assumption is not met, it would be a generalized chi-square distribution [19].

In these cases, critical values can be obtained using a repetitive sampling method with replacement (Monte Carlo simulation). For the simulation, $2^k - 1 - a$ half-normal distributions (truncated normal distribution from the quantile of order 0.5 to the quantile of order 0.9999) are defined. The outcome variable is the sum of squares of the $2^k - 1 - a$ to half-normal distributions. A correlation matrix is defined that has ones in the main diagonal, the values of the significant autocorrelations between the corresponding variables (e.g., with first-order lag corresponds to the contiguous diagonals above and below the main diagonal), and zeros in the remaining cells. See an example of this matrix with three variables (bivariate normality: two variables and their linear combination) with a lag of first order ($h = 1$) and a negative autocorrelation of −0.7.

$$X_1, X_2, \text{ and } X_3 \sim \text{HN } (\sigma = 1).$$

Correlation	X_1	X_2	X_3
X_1	1	−0.7	0
X_2	−0.7	1	−0.7
X_3	0	−0.7	1

Table 2 shows the quantiles for the case of three half-normally distributed variables, as well as the quantiles of the sampling distribution of the Q-statistic when the variables are independent or when they are correlated (with a first-order lag and different values of positive or negative autocorrelation). Compared to the quantiles of a chi-square distribution with three degrees of freedom, the bootstrap-based quantiles with independent variables are slightly higher. However, they increase quite a bit as the variables have higher positive correlations, and decrease quite a bit as the variables have lower negative correlations. These could be cases of bivariate normality testing with two variables, as three simple linear combinations are generated with these two variables ($I = 2^2 - 1 = 3$).

Table 3 shows the bootstrap-based quantiles for seven independent or dependent (with an autocorrelation of −0.5, −0.25, 0.25 or 0.5 for a first-order lag) half-normally distributed variables and the quantiles corresponding to a chi-square distribution with seven degrees of freedom. It is also observed that the bootstrap-based quantiles with independent variables are slightly higher than those of the chi-square distribution with seven degrees of freedom. However, they increase greatly when the variables are positively correlated and decrease greatly when they are negatively correlated. These could be cases of the multivariate normality test with three variables, since seven simple linear combinations are generated ($I = 2^3 - 1 = 7$).

The test of the assumption of serial independence is performed after truncating the standardized values of the log-transformed W or W' statistics, and is therefore performed on the generative sequence of the $2^k - 1$ values z_l or z_l'. These values are reconceptualized as random variables with one-case samples in

Table 2. Bootstrap-based quantiles (Monte Carlo simulation) and quantiles of chi-square distribution with 3 degrees of freedom.

p	Q ind	$\chi^2_{(3)}$	Rel. Q (variables with first-order lagged autocorrelations)									
			−0.7	−0.6	−0.5	−0.4	−0.25	0.25	0.4	0.5	0.6	0.7
0.05	0.367	0.352	0.005	0.072	0.132	0.172	0.286	0.343	0.468	0.445	0.372	0.196
0.10	0.579	0.584	0.008	0.124	0.232	0.326	0.442	0.618	0.701	0.714	0.616	0.380
0.20	0.996	1.005	0.014	0.216	0.401	0.557	0.792	1.116	1.200	1.144	1.063	0.822
0.25	1.198	1.213	0.017	0.261	0.477	0.660	0.980	1.368	1.465	1.411	1.315	1.072
0.30	1.425	1.424	0.020	0.304	0.580	0.757	1.122	1.614	1.739	1.675	1.617	1.292
0.40	1.802	1.869	0.027	0.406	0.748	1.009	1.462	2.141	2.258	2.147	2.143	1.866
0.50	2.271	2.366	0.036	0.502	0.925	1.324	1.868	2.665	2.848	2.755	2.760	2.493
0.60	2.866	2.946	0.045	0.637	1.135	1.637	2.350	3.444	3.521	3.529	3.536	3.369
0.70	3.538	3.665	0.058	0.788	1.428	2.040	2.884	4.224	4.419	4.444	4.345	4.482
0.75	4.060	4.108	0.066	0.900	1.644	2.282	3.209	4.788	5.015	4.993	4.979	5.027
0.80	4.586	4.642	0.077	1.002	1.888	2.601	3.707	5.469	5.754	5.625	5.609	5.821
0.90	6.594	6.251	0.101	1.372	2.448	3.547	4.987	7.275	8.348	8.222	8.094	8.508
0.95	8.322	7.815	0.133	1.723	3.261	4.455	6.388	9.072	10.50	10.25	10.76	11.39
0.975		9.348										
0.99		11.34										
0.995		12.84										

Note: p = quantile order. Three truncated normal distributions (between the quantile of 0.50 order and the quantile of 0.9999 order), *i.e.*, three standard half-normal distributions ($\sigma^2 = 1$), Ind Q = bootstrap-based quantiles for the test statistic or sum of squares of the three half-normally distributed variables when they are independent, and Rel Q when they are correlated (a first-order lag), $\chi^2_{(3)}$ = the pth quantiles of a chi-squared distribution with three degrees of freedom. In each simulation, the number of bootstrap samples was 1000. The sampling method was Latin hypercubes (number of sections = 500), and the correlation type was Spearman. Monte Carlo simulations were performed with XLSTAT version 24 [20].

Table 3. Bootstrap-based quantiles and quantiles of chi-square distribution with 7 degrees of freedom.

p	Ind Q	$\chi^2_{(7)}$	Rel. Q (variables with first-order lagged autocorrelations)			
			−0.5	−0.25	0.25	0.5
0.05	2.233	2.167	0.733	1.438	2.472	2.696
0.10	2.767	2.833	0.977	1.915	3.369	3.381
0.20	3.801	3.822	1.326	2.600	4.618	4.705
0.25	4.155	4.255	1.472	2.886	5.064	5.344

Continued

0.30	4.537	4.671	1.598	3.134	5.616	5.994
0.40	5.451	5.493	1.837	3.601	6.668	7.017
0.50	6.398	6.346	2.114	4.146	7.624	7.989
0.60	7.455	7.283	2.498	4.898	8.712	9.346
0.70	8.484	8.383	2.894	5.674	10.116	11.111
0.75	9.061	9.037	3.138	6.154	10.778	12.132
0.80	9.760	9.803	3.408	6.682	11.647	13.027
0.90	11.957	12.017	4.151	8.140	14.284	16.200
0.95	13.638	14.067	5.060	9.922	16.725	19.647
0.975		16.013				
0.99		18.475				
0.995		20.278				

Note: p = quantile order. Seven truncated normal distributions (between the quantile of 0.50 order and the quantile of 0.9999 order), *i.e.*, seven standard half-normal distributions ($\sigma^2 = 1$), Ind Q = bootstrap-based quantiles for the test statistic or sum of squares of the three half-normally distributed variables when they are independent, and Rel Q when they are correlated (a first-order lag), $\chi^2_{(3)}$ = the pth quantiles of a chi-squared distribution with three degrees of freedom. In each simulation, the number of bootstrap samples was 1000. The sampling method was Latin hypercubes (number of sections = 500), and the correlation type was Spearman. Monte Carlo simulations were performed with XLSTAT version 24 [20].

a cumulative sequence (k individual variables: z_1, z_2, …, z_k, $k(k-1)/2$ sums of two variables, $_kC_3$ sums of three variables, …, a sum of k variables), and this reconceptualization allows testing the hypothesis of independence by means of the Wald-Wolfowitz runs test [21]. If the null hypothesis of independence holds, the chi-square approximation with $2^k - 1 - a$ degrees of freedom could be used. Here the reduction of the degrees of freedom ($-a$) is an operational correction.

Usually, the assumption of random independence within a sequence of data is tested with a first-order lag, as does the Wald-Wolfowitz nonparametric test [20]. Another more comprehensive option to check serial independence is to specify lags from 1 to h and test for significance using the Ljung-Box Q-test [22], as well as the correlogram [23].

When considering that the autocorrelation values are required for the simulation, it is recommended to use the Ljung-Box Q-test; all the more so, when it is the most powerful test of serial dependence in comparative studies [24]. This test has an assumption of bivariate normality in each autocorrelation, so its robust variant can be used, which consists of transforming the variable into ranks, using average ranks in the case of ties [25] [26]. Another option is to use the series without truncation. The maximum number of lags (h) for the Ljung-Box test can be estimated using the Hyndman-Athanasopoulos rule for nonstationary series

[27]: $h = \min(10, n/5)$, where $n = 2^k - 1$. Another option would be the Schwert's rule: $h = 12 \times (n/100)^{0.25}$ [28], widely used in econometrics [29], in turn, a number of lags from 5 to 10 is usually recommended for the correlogram [29].

In case of significance, the analysis is repeated with the simplified series (without zeros) from the ordinary Ljung-Box test (without transforming to ranks). If this second test is significant, the correlogram is used to obtain the significantly non-zero autocorrelations. In case the significance of the full sequence is not confirmed in the simplified sequence (used in the simulation), there is a situation of ambiguity, which is resolved in favor of non-significance. Because the bootstrap-based quantiles with medium or high negative correlations go down a lot with respect to independent samples or up a lot with medium or high positive correlations, it is not recommended to consider only the reduced sequence (without zeros).

2.5. Statistical Decision with an Alpha Significance Level

To make the statistical decision, tested the assumption of independence in the sequence of z_l values, the critical level or probability value of the test statistic value conditional on the null hypothesis of multivariate normality is calculated from the point to infinity (right tail) in a chi-square distribution with $2^k - 1 - a$ degrees of freedom. If $P(\chi^2_{(k-1-a)} \geq q'$ o $q) > \alpha$, the null hypothesis of multivariate normality holds, since the test statistic value falls within the acceptance region: q' or $q \leq {}_{1-\alpha}\chi^2_{(k-1-a)}$. Conversely, if $P(\chi^2_{(k-1-a)} \geq q'$ o $q) \leq \alpha$, the null hypothesis of multivariate normality is rejected, since the test statistic value falls within the rejection region: q' o $q > {}_{1-\alpha}\chi^2_{(k-1-a)}$.

If the assumption does not hold, a simulation would be used to obtain the quantiles or critical values (percentiles) by generating at least 1000 bootstrap random samples. The sampling distribution of the $2^k - 1 - a$ variables (non-negative z_l values) would be truncated standard normal distribution (from quartile of 0.5 order to quartile of 0.9999 order), *i.e.* standard half-normal distribution (with scale parameter $\sigma = 1$), and the outcome variable would be the sum of squares of the $2^k - 1 - a$ half-normally distributed variables. The matrix is defined with the significant autocorrelations (*ar*). If q' or $q > P_{95}$ (bootstrap-based 95th percentile), H_0 is rejected. It should be noted that the discrepancy between the bootstrap-based quantiles (from a repeated sampling with replacement or Monte Carlo simulation) and quantiles of a chi-square distribution is not large in the case of the number of variables is high and the correlations low (*ar* < 0.30). The bootstrap-based quantiles are higher if the correlations are positive and lower if they are negative, so that chi-square quantiles of order 0.975 (*ar* = 0.25), 0.99 (*ar* = 0.3) and 0.995 (*ar* = 0.4) or 0.9 (*ar* = −0.25), 0.85 (*ar* = −0.3), and 0.8 (*ar* = −0.4) could also be used instead of 0.95 in case the simulation fails, because of its complexity.

2.6. Effect Size

The effect size can be calculated using the squared eta coefficient (η^2) or its

square root which is the eta coefficient (η). Both vary from 0 (no effect or independence) to 1 (total effect or dependence). Values of η^2 less than 0.01 are usually considered trivial. Cohen [30] suggests interpreting η^2 values between 0.01 and 0.059 as a small effect size, between 0.06 and 0.139 medium, and greater than or equal to 0.14 large for analysis of variance. For multiple linear regression, the cut-off points suggested by Cohen [30] are: 0.02 (small), 0.13 (medium), and 0.26 (large). A larger effect implies a larger deviation from the multivariate normality model in the test being proposed.

$$\eta^2 = \frac{Q'}{n \times gl} = \frac{Q'}{n\left(2^k - 1 - a\right)} \quad \text{o} \quad \eta^2 = \frac{Q}{n\left(2^k - 1 - a\right)}$$

2.7. A Posteriori Type II Error and Statistical Power

Type II error (β) and statistical power ($\phi = 1 - \beta$) are calculated using the cumulative distribution function of the noncentral chi-square distribution ($NC\chi^2$). Its degrees of freedom are the k-th power of 2 subtracting 1 and the number of null values, $v = 2^k - 1 - a$, and the non-centrality parameter is the squared eta coefficient multiplied by the sample size and degrees of freedom, $\lambda = \eta^2 \times n \times (2^k - 1 - a) = q$. The distribution function is evaluated at the critical value of the statistic Q or Q': $_{1-a}\chi^2[2^k - 1 - a]$, i.e., the quantile of order $1 - \alpha$ of a chi-squared distribution with $2^k - 1 - a$ degrees of freedom.

$$\beta = NC\chi^2_{\left(v=2^k-1-a,\,\lambda=\eta^2\times n\times\left(2^k-1-a\right)=q\right)}\left(_{1-\alpha}\chi^2_{2^k-1-a}\right)$$

$$\phi = 1 - \beta$$

$\beta = P(\text{hold } H_0 | H_1 \text{ true})$ = the type II error or false negative error = the probability of holding the null hypothesis conditional on a true alternative hypothesis. When the null hypothesis is rejected due to a probability value less than the significance level, $p < \alpha$, the probability β should be less than 0.5 and preferably equal to or less than 0.2, evidencing low probability of a mistaken rejection. When the null hypothesis holds due to a probability value greater than or equal to the significance level α, $p \geq \alpha$, the probability β should be greater than 0.5, indicating the low probability of the alternative hypothesis.

$\phi = P(\text{reject } H_0 | H_1 \text{ true})$ = statistical power or success in rejecting the null hypothesis = probability of rejecting the null hypothesis conditional on a true alternative hypothesis. When the null hypothesis is rejected due to a probability value less than the significance level, $p < \alpha$, the statistical power should be greater than 0.5 and preferably equal to or greater than 0.8 (strong power) or 0.9 (very strong power). When the null hypothesis holds due to a probability value greater than or equal to the significance level, $p \geq \alpha$, the statistical power should be less than 0.5, revealing the low probability of the alternative hypothesis.

$\alpha = P(\text{reject false } H_0 | H_1)$ = type I error or false positive error, which is also called significance level = the probability of rejecting the null hypothesis conditional on a false alternative hypothesis and hence true null hypothesis. It is fixed a priori, usually with a value of 0.05. With small samples, it can be raised to 0.1

and with very large samples can be lowered to 0.01.

$1 - \alpha = P(\text{hold } H_0 | H_1 \text{ false}) = $ success in holding the null hypothesis or confidence level = the probability of holding the null hypothesis conditional on a false alternative hypothesis and hence true null hypothesis. If $\alpha = 0.05$, then $1 - \alpha = 0.95$.

2.8. Example of Calculation of the Proposed Test

Let be a random sample of 20 participants who were measured on four different variables (Table 4). Check whether the sample was drawn from a population with multivariate normal distribution of unknown parameters μ and Σ.

$$H_0 : \vec{x} = \begin{pmatrix} X_1 & X_2 & X_3 & X_4 \end{pmatrix} \sim N(\vec{\mu}, \Sigma)$$

$$H_1 : \vec{x} = \begin{pmatrix} X_1 & X_2 & X_3 & X_4 \end{pmatrix} \nsim N(\vec{\mu}, \Sigma)$$

The 15 unweighted linear combinations (by simple summation) of the four variables are calculated ($l = 2^4 - 1 = 15$). The 20 data of each linear combination of variables are sorted in ascending order (empirical quantiles). These data are

Table 4. Data of the four variables in their random order.

Participant	X_1	X_2	X_3	X_4
1	−0.354	−0.787	−0.360	−0.281
2	0.008	−0.573	−1.045	0.627
3	0.107	0.211	0.372	−0.473
4	−0.142	1.940	−0.085	0.973
5	−0.222	−0.081	1.395	−0.616
6	1.629	0.540	0.215	−0.298
7	0.020	0.225	−0.146	0.517
8	0.482	−0.334	−0.032	0.001
9	−0.487	1.702	−0.713	1.330
10	−0.249	0.645	0.744	0.047
11	−0.508	0.560	−1.148	0.473
12	−1.756	−1.081	−1.560	1.406
13	−1.091	0.288	0.588	0.078
14	1.916	0.573	0.690	1.305
15	1.493	−1.180	−0.288	0.197
16	−0.337	−0.357	0.032	0.199
17	0.559	−0.203	1.104	0.232
18	0.463	−0.717	0.699	0.768
19	1.633	−0.577	0.317	−0.118
20	−0.670	1.591	−0.585	0.008

assigned ranks i from 1 to 20. The orders of the theoretical quantiles, denoted by p_i, are obtained as a function of the order i, $p_i = (i - 0.375)/20.25$. The theoretical quantiles, denoted by m_i, are computed using the Probit function or quantile function of the standard normal distribution, $m_i = \Phi^{-1}(p_i)$. Finally, the standardized and normalized expected values, denoted by a_i, are obtained using the Royston's formulas [3], as can be seen in Table 5.

On the one hand, the correlations between the empirical quantiles (of the 15 linear combinations) and the standardized and normalized theoretical quantiles a_i are calculated. It should be recalled that theoretical quantiles are the same for all 15 linear combinations. These squared correlation coefficients constitute the W test statistics of the Shapiro-Wilk univariate normality test [1]. A logarithmic transformation is applied to the W statistics, the mean and standard deviation of the transformation are calculated as a function of the sample size ($n = 20$), and the log-transformed statistics are standardized following the Royston's procedure [3]. See Table 6.

Table 5. Linear combinations (of one and two variables) among the 4 variables with their values sorted in ascending order or empirical quantiles and expected quantiles.

i	X_1	X_2	X_3	X_4	$X_1 + X_2$	...	$X_1 + X_2 + X_3 + X_4$	p_i	m_i	a_i
1	−1.756	−1.180	−1.560	−0.616	−2.837		−2.990	0.031	−1.868	−0.473
2	−1.091	−1.081	−1.148	−0.473	−1.141		−1.781	0.080	−1.403	−0.322
3	−0.670	−0.787	−1.045	−0.298	−0.803		−0.983	0.130	−1.128	−0.256
4	−0.508	−0.717	−0.713	−0.281	−0.694		−0.622	0.179	−0.919	−0.208
5	−0.487	−0.577	−0.585	−0.118	−0.565		−0.462	0.228	−0.744	−0.169
6	−0.354	−0.573	−0.360	0.001	−0.302		−0.137	0.278	−0.589	−0.134
7	−0.337	−0.357	−0.288	0.008	−0.254		0.116	0.327	−0.448	−0.101
8	−0.249	−0.334	−0.146	0.047	0.052		0.217	0.377	−0.315	−0.071
9	−0.222	−0.203	−0.085	0.078	0.148		0.223	0.426	−0.187	−0.042
10	−0.142	−0.081	−0.032	0.197	0.246		0.344	0.475	−0.062	−0.014
11	0.008	0.211	0.032	0.199	0.313		0.476	0.525	0.062	0.014
12	0.020	0.225	0.215	0.232	0.318		0.617	0.574	0.187	0.042
13	0.107	0.288	0.317	0.473	0.356		1.187	0.623	0.315	0.071
14	0.463	0.540	0.372	0.517	0.396		1.212	0.673	0.448	0.101
15	0.482	0.560	0.588	0.627	0.921		1.255	0.722	0.589	0.134
16	0.559	0.573	0.690	0.768	1.056		1.693	0.772	0.744	0.169
17	1.493	0.645	0.699	0.973	1.214		1.831	0.821	0.919	0.208
18	1.629	1.591	0.744	1.305	1.798		2.086	0.870	1.128	0.256
19	1.633	1.702	1.104	1.330	2.169		2.686	0.920	1.403	0.322
20	1.916	1.940	1.395	1.406	2.489		4.484	0.969	1.868	0.473

Note: i = the order of the data when sorted in ascending order, $p_i = (i - 0.375)/20.25$ = the order of the theoretical quantile [16], $m_i = \Phi^{-1}(p_i)$ = theoretical quantile under normal distribution [3] [4], a_i = standardized and normalized theoretical quantile [3].

Table 6. Statistics W and W' with their logarithmic transformation, standardization and truncation.

	I	w_I	$\ln(1-w_I)$	$z_{\ln(1-W_I)}$	z_I	w'_I	$\ln(1-w')$	$z_{\ln(1-W'_I)}$	z'_I
1	X_1	0.939	−2.793	0.749	0.749	0.939	−2.801	0.839	0.839
2	X_2	0.941	−2.832	0.669	0.669	0.950	−2.986	0.508	0.508
3	X_3	0.986	−4.253	−2.194	0	0.989	−4.500	−2.203	0
4	X_4	0.946	−2.927	0.478	0.478	0.958	−3.169	0.181	0.181
5	$X_1 + X_2$	0.962	−3.276	−0.225	0	0.951	−3.023	0.442	0.442
6	$X_1 + X_3$	0.975	−3.699	−1.079	0	0.973	−3.607	−0.604	0
7	$X_1 + X_4$	0.940	−2.809	0.716	0.716	0.936	−2.754	0.924	0.924
8	$X_2 + X_3$	0.943	−2.859	0.615	0.615	0.945	−2.903	0.658	0.658
9	$X_2 + X_4$	0.907	−2.376	1.588	1.588	0.913	−2.437	1.492	1.492
10	$X_3 + X_4$	0.951	−3.024	0.282	0.282	0.958	−3.163	0.192	0.192
11	$X_1 + X_2 + X_3$	0.941	−2.825	0.683	0.683	0.924	−2.576	1.242	1.242
12	$X_1 + X_2 + X_4$	0.939	−2.790	0.754	0.754	0.939	−2.795	0.850	0.850
13	$X_1 + X_3 + X_4$	0.966	−3.372	−0.419	0	0.959	−3.199	0.126	0.126
14	$X_2 + X_3 + X_4$	0.958	−3.166	−0.004	0	0.969	−3.475	−0.368	0
15	$X_1 + X_2 + X_3 + X_4$	0.979	−3.866	−1.415	0	0.964	−3.332	−0.111	0

Note: I = the order of the simple linear combination in its generative sequence, w_I = the Shapiro-Wilk W-test statistic of the lth combination, $\ln(1-w_I)$ = the log-transformation of w_I statistic, $z_{\ln(1-W_I)}$ = the standardized value of the log-transformation of w_I statistic, z_I = the standardized and truncated value of the log-transformation of w_I statistic ($z_I = 0$, if $z_{\ln(1-W_I)} < 0$), w'_I = the Shapiro-Francia W' test statistic of the lth combination, $\ln(1-w'_I)$ = the log-transformation of w'_I statistic, $z_{\ln(1-W'_I)}$ = standardized value of the log-transformation of w'_I statistic, z'_I = the standardized and truncated value of the log-transformation of w'_I statistic ($z'_I = 0$, if $z_{\ln(1-W'_I)} < 0$).

$$w_1 = r^2_{X_1 a} = \frac{\left(\sum_{i=1}^{20}(x_{i1}-\bar{x}_1)a_i\right)^2}{\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2} = \frac{4.0262^2}{17.2680} = 0.9387$$

$$\mu_{\ln(1-W)|n=20} = -1.5861 - 0.31082\ln(n) - 0.083751\left[\ln(n)\right]^2$$
$$+ 0.0038915\left[\ln(n)\right]^3$$
$$= -1.5861 - 0.31082\ln(20) - 0.083751\left[\ln(20)\right]^2$$
$$+ 0.0038915\left[\ln(20)\right]^3$$
$$= -3.1642$$

$$\sigma_{\ln(1-W)|n=20} = e^{-0.4803 - 0.082676\ln(n) + 0.0030302\left[\ln(n)\right]^2}$$
$$= e^{-0.4803 - 0.082676\ln(20) + 0.0030302\left[\ln(20)\right]^2}$$
$$= 0.4962$$

$$Z_{\ln(1-W_1)} = \frac{\ln(1-W_1) - \mu_{\ln(1-W)}}{\sigma_{\ln(1-W)}} = \frac{\ln(1-0.9387) - (-3.1642)}{0.4962}$$

$$= \frac{-2.7927 + 3.1642}{0.4962} = 0.7488$$

On the other hand, the correlations between the empirical quantiles of the 15 linear combinations and the theoretical quantiles m_i that are common to the 15 combinations are calculated. These squared correlation coefficients constitute the W' test statistics of the Shapiro-Francia univariate normality test [2]. A logarithmic transformation is applied to the W' statistics, the mean and standard deviation of the transformation are calculated as a function of the sample size (n = 20), and the log-transformed statistics are standardized following the Royston's procedure [4]. Finally, the negative values of $\ln(1-W_l)$ and $\ln(1-W_l')$ are cancelled, which affects six values in the first case and four in the second, resulting in the sequences z_l and z_l'. See **Table 6**.

$$w_1' = r_{X_1 m}^2 = \frac{\left[\sum_{i=1}^{20}(x_{i1} - \bar{x}_1)m_i\right]^2}{\sum_{i=1}^{20}(x_{i1} - \bar{x}_1)^2 \sum_{i=1}^{20}m_i^2} = \frac{16.9117^2}{17.2680 \times 17.6336} = \frac{286.0060}{304.4959} = 0.9393$$

$$u = \ln\left[\ln(n)\right] - \ln(n) = \ln\left[\ln(20)\right] - \ln(20) = -1.8985$$

$$\mu_{\ln(1-W')} = 1.0528u - 1.2725 = 1.0528 \times (-1.8985) - 1.2725 = -3.26996$$

$$v = \ln(\ln(n)) + 2/\ln(n) = \ln(\ln(20)) + 2/\ln(20) = 1.7648$$

$$\sigma_{\ln(1-W')} = -0.26758v + 1.0308 = -0.26758 \times 1.7648 + 1.0308 = 0.5586$$

$$Z_{\ln(1-w_1')} = \frac{\ln(1-w_1') - \mu_{\ln(1-W')}}{\sigma_{\ln(1-W')}} = \frac{\ln(1-0.9393) - (-3.26996)}{0.5586}$$

$$= \frac{-2.8014 + 3.26996}{0.5586} = 0.8388$$

The assumption of independence between the 15 values of z_l (standardized and truncated values of the log-transformed W statistics) and z_l' (standardized and truncated values of the log-transformed W' statistics) is tested using the Wald-Wolfowitz runs test from the exact probabilities at a 10% significance level due to the small sizes of the sequences (15 elements). The hypothesis holds at the 10% significance level in a two-tailed test for both sequences (in its generative order). The exact probabilities are the same for both sequence of values z_l and z_l', since the number of runs (r = 8), number of values z_l or z_l' less than their median (n_o = 7) and greater than or equal to their median (n_1 = 8) coincide: point probability, $P(R = 8) = 0.2176$, probability to the left tail, $P[R \leq 8 = Mdn(R)] = 0.5136$, probability to the other tail, $P(R > 8) = 0.4864$, and probability to two tails, $P(R \leq 8) + P(R > 8) = 0.5136 + 0.4864 = 1 > \alpha = 0.10$.

In turn, the independence assumption is tested using the Ljung-Box test. The lag order was determined using the Hyndman-Athanasopoulos criterion for non-stationary series: h = min(10, $n/5$) = min(10, 15/5) = 3. The test is not sig-

nificant at the 10% significance level for either sequence with a third-order lag.

A visual inspection of the two correlograms with a maximum lag of 7 (Schwert's criterion) reveals no significant autocorrelation, since the bars lie in the space between the upper and lower limits of 90% confidence intervals of null autocorrelations with lags from 1 to 7. See **Figure 1**.

$$h = 12 \times (n/100)^{1/4} = 12 \times (15/100)^{1/4} = 7.46 = 7$$

Lung-Box test applied to the $z_{\ln(1-W)}$ sequence (from the Shapiro-Wilk W statistics)

$$Q_{LB} = n(n+2)\sum_{i=1}^{h} \frac{ar_i^2}{n-i} = 15 \times 17 \times \left(\frac{-0.0391^2}{14} + \frac{0.0457^2}{13} + \frac{0.0989^2}{12} \right) = 0.0279$$

$$q_{LB} = 0.0279 < {}_{1-\alpha}\chi_h^2 = {}_{0.90}\chi_3^2 = 6.2514$$

y $P\left(\chi_3^2 \geq q_{LB} = 0.0279\right) = 0.9988 > \alpha = 0.10$, se mantiene H_0: $\rho_1 = \rho_2 = \rho_3 = 0$.
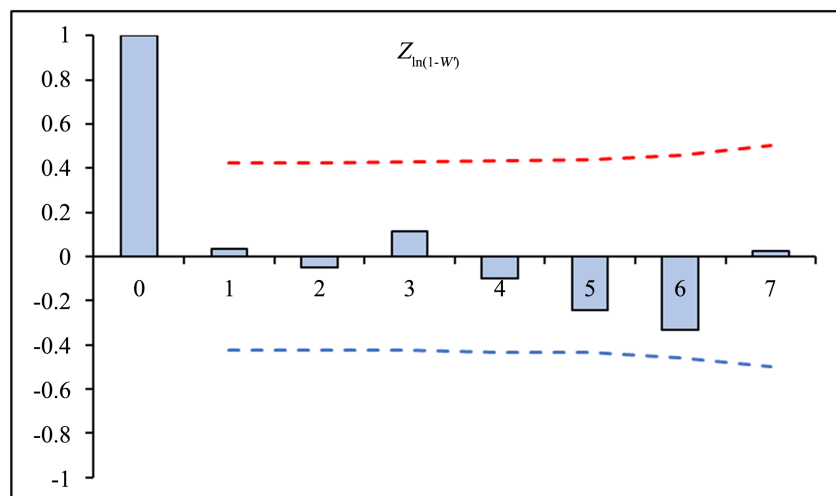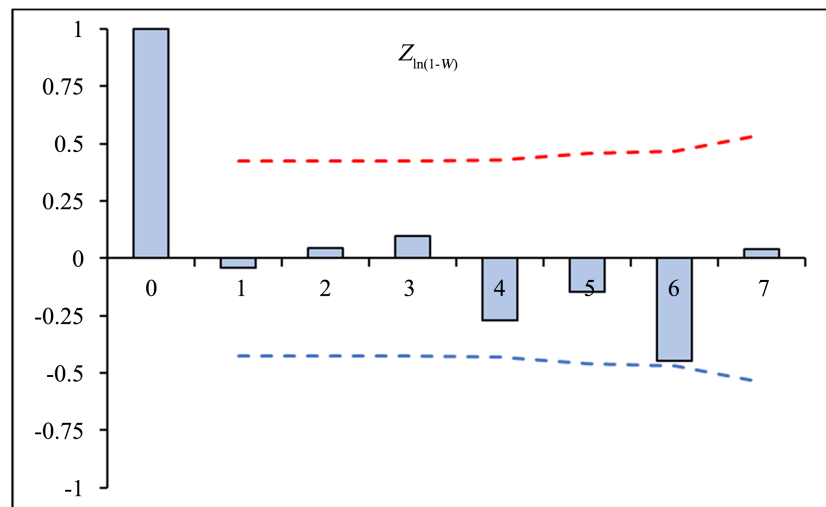

Figure 1. Correlograms with interpretation limits with a 90% confidence level.

Lung-Box test applied to the $z_{\ln(1-W')}$ sequence (from the Shapiro-Francia W' statistics).

$$Q_{LB} = n(n+2)\sum_{i=1}^{h}\frac{ar_i^2}{n-i} = 15\times17\times\left(\frac{0.0357^2}{14}+\frac{-0.0489^2}{13}+\frac{0.1135^2}{12}\right) = 0.0232$$

$$q_{LB} = 0.0232 < {}_{1-\alpha}\chi_h^2 = {}_{0.90}\chi_3^2 = 6.2514$$

y $P\left(\chi_3^2 \geq q_{LB} = 0.0232\right) = 0.9991 > \alpha = 0.10$, se mantiene H_0: $\rho_1 = \rho_2 = \rho_3 = 0$.

With the robust Ljung-Box test [25] [26], the null hypothesis of independence also holds for the sequence z_l ($Q_r = 0.338$, $p = 0.953$) and z_l' ($Q_r = 0.620$, $p = 0.892$) with third-order lag.

Next, the sums of squares of z_l and z_l' are performed, yielding the test statistics q and q', respectively. Since the assumption of independence holds, these two sums of squares follow a chi-square distribution. The degrees of freedom are 9 for Q-test (from the z_l statistics) and 11 for Q'-test (from the z_l' statistics). Since the values of the test statistics q and q' are less than their corresponding critical values and their probability values under the null hypothesis of multivariate normality are greater than the 5% significance level, the null hypothesis is maintained with both versions of the proposed test.

$$Q = \sum_{l=1}^{2^k-1} z_l^2 = \sum_{l=1}^{15} z_l^2$$
$$= 0.7488^2 + 0.6686^2 + 0 + 0.4778^2 + 0 + 0 + 0.7162^2 + 0.6148^2$$
$$+ 1.5879^2 + 0.2825^2 + 0.6828^2 + 0.7544^2 + 0 + 0 + 0$$
$$= 5.7636$$

$$df = 2^k - 1 - a = 2^4 - 1 - 6 = 16 - 1 - 6 = 9$$

$$q = 5.7636 < {}_{1-\alpha}\chi_{2^k-1-a}^2 = {}_{0.95}\chi_9^2 = 16.9190$$

$$P\left(\chi_9^2 \geq q = 5.7636\right) = 0.7633 > \alpha = 0.05$$

$$Q' = \sum_{l=1}^{2^k-1}(z_l')^2 = \sum_{l=1}^{15}(z_l')^2$$
$$= 0.8388^2 + 0.5081^2 + 0 + 0.1805^2 + 0.4418^2 + 0 + 0.9239^2 + 0.6576^2$$
$$+ 1.4918^2 + 0.1916^2 + 1.2416^2 + 0.8499^2 + 0.1262^2 + 0 + 0$$
$$= 7.0174$$

$$df = 2^k - 1 - a = 2^4 - 1 - 4 = 16 - 1 = 11$$

$$q' = 7.0174 < {}_{1-\alpha}\chi_{2^k-1-4}^2 = {}_{0.95}\chi_{11}^2 = 19.6751$$

$$P\left(\chi_{11}^2 \geq q' = 7.0174\right) = 0.7977 > \alpha = 0.05$$

The effect size is calculated, which is small in both versions of the proposed test.

$$\eta^2 = \frac{Q}{n\times gl} = \frac{Q}{n\left(2^k-1-a\right)} = \frac{5.7636}{20\times9} = 0.0320$$

$$\eta^2 = \frac{Q'}{n \times gl} = \frac{Q'}{n\left(2^k - 1 - a\right)} = \frac{7.0174}{20 \times 11} = 0.0319$$

The type II error is very high ($\beta > 0.5$) and the statistical power very low ($\phi = 1 - \beta < 0.5$), which indicates that the alternative hypothesis is very unlikely and the null hypothesis of multivariate normality should be maintained with both versions of the proposed test. The version with the Shapiro-Wilk W statistic [3] seems slightly better than with the Shapiro-Francia W' statistic [4] in terms of type II error and statistical power.

Calculation of type II error and statistical power for Q-test (from Shapiro-Wilk W statistics).

$$\beta = NC\chi^2_{\left(\upsilon = 2^k - 1 - a, \lambda = \eta^2 \times n \times \left(2^k - 1 - a\right) = q\right)}\left({}_{1-\alpha}\chi^2_{2^k - 1 - a}\right)$$

$$\beta = NC\chi^2_{(\upsilon = 9, \lambda = 5.7636)}\left({}_{0.95}\chi^2_9 = 16.9190\right) = 0.6750$$

$$\phi = 1 - \beta = 1 - 0.6750 = 0.3250$$

Calculation of type II error and statistical power for Q'-test (from Shapiro-Francia W' statistics).

$$\beta = NC\chi^2_{\left(\upsilon = 2^k - 1, \lambda = \eta^2 \times n \times (2^k - 1) = q'\right)}\left({}_{1-\alpha}\chi^2_{2^k - 1}\right)$$

$$\beta = NC\chi^2_{(\upsilon = 11, \lambda = 7.0174)}\left({}_{0.95}\chi^2_{11} = 16.9190\right) = 0.6360$$

$$\phi = 1 - \beta = 1 - 0.6360 = 0.3640$$

3. Method

3.1. Sample Generation and Statistical Analysis

In the present simulation study, only samples of 20 elements for four variables were analyzed, so it is a pilot study of the proposal. The small size of 20 was chosen as it is the minimum recommended for normality tests and with which the statistical power and discriminative capacity of the tests can be more compromised. Four variables were chosen as this is an easily manageable small number.

Fifty samples of 20 elements and 4 variables were generated. On the one hand, 40 samples were created. Four random variables with standard continuous uniform distribution U[0, 1] were used as a starting point to obtain variables with a standard normal distribution: $x_i = \Phi^{-1}(p_i)$; with a good distributional convergence to normality: chi-square with 100 degrees of freedom, ${}_{p_i}\chi^2_{100}$, Student's t with 100 degrees of freedom, ${}_{p_i}t_{100}$, and binomial B(p_i, 0.5, 20); with a distribution close to normality: standard logistic distribution: $x_i = \ln[p_i/(1 - p_i)]$; and with distributions far from normality: exponential with parameter $\lambda = 1$: $x_i = \ln(p_i)$, standard Cauchy distribution: $x_i = \tan[\pi \times (p_i - 0.5)]$, standard log normal distribution: $\exp(\Phi^{-1}(p_i))$, standard Laplace distribution: $x_i = -1 \times \text{sign}(p_i - 0.5) \times \ln(1 - 2 \times |p_i - 0.5|)$, and binomial: $x_i = B(p_i, 0.1, 10)$. Samples with four correlated normally distributed variables were obtained using Cholesky decomposi-

tion of covariance matrices with moderate to high correlations. In addition, 10 more samples were created. Eight standard continuous uniform variables were used to generate eight independent normal variables (Z), and from their transformation, independent non-normal distributions were obtained: inverse normal ($Z^{-1} = 1/Z$), chi-square with one degree of freedom (Z^2), chi-square with two degrees of freedom ($Z_1^2 + Z_2^2$), chi-square four degrees of freedom ($Z_1^2 + Z_2^2 + Z_3^2 + Z_4^2$), and Cauchy distribution (Z_1/Z_2).

It was considered that the normality hypothesis should hold in the following situations (10 samples, 20%):

- Seven samples with normal variables (3 samples with the four independent variables and 4 samples with the four correlated variables).
- Three samples with a good convergence to multivariate normality: four variables following a Student's t-distribution with 100 degrees of freedom, four variables following a chi-square distribution with 100 degrees of freedom, and four variables following a binomial distribution B($n = 20$, $p = 0.5$).

Conversely, it should be rejected in the following situations (40 samples, 80%): N = normal, E = exponential, LogN = log normal, Logist = logistic, Lap = Laplace, C = Cauchy, B(n, p) = binomial with n independent trials and constant probability of success p, $N^{-1} = 1/N$ = inverse normal and $\chi^2_{(v)}$ = chi-square with v degrees of freedom.

- Three variables with normal distribution and one without normal distribution (the preceding number indicates the number of variables, but if there is only one variable, the number is omitted): 3NE, 3NC, 3NLogN, 3NN^{-1}, 3N$\chi^2_{(2)}$, 3N$\chi^2_{(1)}$, and 3NCauchy.
- Two independent variables with normal distribution and two without normal distribution (if the subscripts match, they are related, and if they do not match or have no subscripts, they are independent): 2NE$_1$C$_1$, 2NE$_1$LogN$_1$, 2NC$_1$LogN$_1$, 2NE$_2$C$_2$, 2NE$_2$Log$_2$, 2NC$_2$LogN$_2$, 2NE$_3$C$_3$, 2NE$_3$LogN$_3$. 2NC$_3$LogN$_3$, 2NE$_4$C$_4$, 2NE$_4$LogN$_4$, 2NC$_4$LogN$_4$, 2NE$_1$C$_3$, 2NE$_1$LogN$_3$, 2NC$_1$LogN$_3$, 2NE$_2$C$_1$, 2NE$_2$LogN$_1$, 2NC$_2$LogN$_1$, 2N2$\chi^2_{(2)}$, 2N2$\chi^2_{(4)}$, 2N2$\chi^2_{(1)}$, and 2N2C.
- One variable with normal distribution and three without normal distribution: N$_1$E$_1$C$_1$LogN$_1$, N$_2$E$_2$C$_2$LogN$_2$, N$_3$E$_3$C$_3$LogN$_3$, N$_4$E$_4$C$_4$LogN$_4$, and N$_1$E$_2$C$_3$LogN$_4$.
- Four independent variables with no normal distribution: 4E, 4C, 4LogN, 4Logist, 4Lap, and 4B ($n = 10$, $p = 0.1$).

On the one hand, the probability of the correct decision conditional on the alternative hypothesis is calculated. When the null hypothesis must be retained, this probability is the type II error or probability β, and when it must be rejected, it is the statistical power or complement of the beta probability: $\phi = 1 - \beta$. The comparison of measures of central tendency in β or ϕ values among six samples is performed using the Friedman's test [31]. Effect size is calculated using the Kendall's W coefficient of concordance [32]. Pairwise comparisons are performed using the Wilcoxon's signed-rank test [33], estimating the effect size by the Rosenthal's r coefficient [34]. The Sidak's correction is applied with the Holm's procedure to control for the family error rate [35] [36]. These compari-

sons are made with the 50 multivariate samples (β and ϕ), with 40 multivariate samples that deviate from multivariate normality (β), and with the 10 multivariate samples that follow or have a with good convergence to multivariate normality (ϕ).

On the other hand, the frequency of successes among the six tests was compared using the Cochran's Q test [37] with the 50 multivariate samples. Effect size was estimated by the eta-squared coefficient [38]. Pairwise comparisons were performed using the McNemar's test [39], and the effect size was calculated by the Cohen's q statistic (1988). The Sidak's correction with the Holm's procedure was applied to control the error rate by family [35] [36]. In addition, 2×2 tables were computed for each of the six tests with respect to the decision on null hypothesis. From these tables, point and interval estimates of sensitivity, specificity and efficacy were calculated for each of the 50 multivariate samples generated, using the Wilson's confidence interval score with the Newcombe's continuity correction [40] [41]. Pairwise comparisons between these correlated proportions were performed using the McNemar's Z-test [39].

A final analysis involved creating a variable of ordered categories by classifying the 50 multivariate samples generated by their deviation from multivariate normality. Five levels of deviation were defined. The ordinal variable was correlated with the critical level or probability value of each of the six multivariate normality tests, that is, with the probability associated with a critical region bounded by the observed value of the test statistic under the assumption that the null hypothesis is true within a known distribution (normal for the multivariate runs test and chi-square for the other tests). The more negative this correlation is (closer to −1), the better the multivariate normality test. Correlations were calculated by the Spearman's rank coefficient [42]. The confidence interval was estimated with the Fisher's transformation [43] and the Bonett-Wright standard error [44]. Comparisons were made using the Steiger's Z-test [45], following the suggestion of Myers and Sirois for the Spearman's correlation [46].

3.2. Multivariate Normality Tests with Which the New Proposal Is Compared

The calculation of the test statistic, the probability value for the statistical decision at the α level of significance, the effect size, the type II error, and the statistical power of the four tests with which the new proposal is compared are shown below. It starts with the omnibus test based on the multivariate skewness and kurtosis [5] [6] [7]. It continues with the runs Z-test for multivariate normality [8] [9], and ends with Royston's test [10] based on the H statistic (with the transformation and standardization of the W statistic of Royston [3]) and the H' statistic (with the transformation and standardization of the W' statistic of Royston [4]).

3.2.1. Mardia's Omnibus Test of Multivariate Normality
The Mardia's test starts from the square matrix of order n of standardized dis-

tances or moments, denoted by M. This is obtained by the product among the rectangular matrix of order $n \times k$ of differential scores (D), the square matrix of order k of the inverse of sample covariance matrix (S^{-1}), and the rectangular matrix of order $k \times n$ of the transpose of the matrix of the differential scores (D^T), where n is the number of participants and k the number of variables.

$$D = X - \vec{1}^T \vec{x} = \begin{pmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nk} \end{pmatrix} - \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \begin{pmatrix} \overline{x}_1 & \cdots & \overline{x}_k \end{pmatrix}$$

$$S = \frac{1}{n-1} D^T D = \begin{pmatrix} s_{x_1}^2 & \cdots & s_{x_1 x_k} \\ \vdots & \ddots & \vdots \\ s_{x_k x_1} & \cdots & s_{x_k}^2 \end{pmatrix}$$

$$M = D \times S^{-1} \times D^T$$

$$M = \begin{pmatrix} x_{11} - \overline{x}_1 & \cdots & x_{1k} - \overline{x}_k \\ \vdots & \ddots & \vdots \\ x_{n1} - \overline{x}_1 & \cdots & x_{nk} - \overline{x}_k \end{pmatrix} \begin{pmatrix} s_{x_1}^2 & \cdots & s_{x_1 x_k} \\ \vdots & \ddots & \vdots \\ s_{x_k x_1} & \cdots & s_{x_k}^2 \end{pmatrix}^{-1} \begin{pmatrix} x_{11} - \overline{x}_1 & \cdots & x_{n1} - \overline{x}_1 \\ \vdots & \ddots & \vdots \\ x_{1k} - \overline{x}_k & \cdots & x_{nk} - \overline{x}_k \end{pmatrix}$$

$$= \begin{pmatrix} m_{11} & \cdots & m_{1n} \\ \vdots & \ddots & \vdots \\ m_{n1} & \cdots & m_{nn} \end{pmatrix}$$

Mardia's multivariate skewness (b_{1M}) with a correction for sampling bias (b_{1M}^*).

$$b_{1M} = \frac{\sum_{i=1}^n \sum_{j=1}^n m_{ij}^3}{n^2}$$

$$b_{1M}^* = \left(\frac{n}{n-1}\right)^3 b_{1M} = \left(\frac{n}{n-1}\right)^3 \frac{\sum_{i=1}^n \sum_{j=1}^n m_{ij}^3}{n^2}$$

$$Q = \frac{n}{6} b_{1M}^*$$

$$Q_c \sim \chi^2_{\frac{k(k+1)(k+2)}{6}}$$

Mardia's multivariate kurtosis (b_{2M}) with a correction for sampling bias (b_{2M}^*).

$$b_{2M} = \frac{\sum_{i=j=1}^n m_{ij}^2}{n}$$

$$b_{2M}^* = \left(\frac{n}{n-1}\right)^2 b_{2M} = \left(\frac{n}{n-1}\right)^2 \frac{\sum_{i=j=1}^n m_{ij}^2}{n}$$

$$Z = \frac{b_{2M}^* - E(b_{2M})}{SD(b_{2M})} = \frac{b_{2M}^* - k(k+2)}{\sqrt{\frac{8k(k+2)}{n}}} \sim N(0,1)$$

Test statistics and sampling distribution

$$K_{MVN}^2 = Q + Z^2$$

$$K_{MVN}^2 \sim \chi^2_{\left(\frac{k(k+1)(k+2)}{6}+1\right)}$$

Statistical decision under the null hypothesis of multivariate normality with a given level of significance (alpha).

$$P\left(\chi^2_{\left(\frac{k(k+1)(k+2)}{6}+1\right)} \geq K_{MVN}^2\right) \geq \alpha \text{ , H}_0 \text{ is accepted; and } < \alpha \text{ , H}_0 \text{ is rejected.}$$

The effect size estimated using the eta-squared coefficient.

$$\eta^2 = \frac{K_{MVN}^2}{n \times df} = \frac{K_{MVN}^2}{n\left(\frac{k(k+1)(k+2)}{6}+1\right)}$$

Type II error and statistical power calculated using the cumulative distribution function of a noncentral chi-squared distribution.

$$\beta = NC\chi^2_{df=1+\frac{k(k+1)(k+2)}{6}, NCP=\eta^2 \times n \times df = K_{MVN}^2}\left(1-\alpha\,\chi^2_{\frac{k(k+1)(k+2)}{6}+1}\right)$$

$$\phi = 1 - \beta$$

3.2.2. Royston's Multivariate Normality Test and Its Two Variants

To obtain the test statistic H (from Shapiro-Wilk W statistics) or H' (from Shapiro-Francia W' statistics), the following formulas are used [10]:

$$H = \frac{e\sum_{j=1}^{k}\psi_j}{k}$$

$$\psi_j = \left[\Phi^{-1}\left(\frac{\Phi(-z_j)}{z}\right)\right]^2 = \left[\Phi^{-1}\left(0.5 \times \Phi(-z_j)\right)\right]^2, j = 1, 2, \cdots, k$$

Φ^{-1} = the probit function or quantile function associated with a standard normal distribution, and Φ = the cumulative distribution function of a standard normal variable.

$$Z_j = \frac{\ln(1-W_j) - \mu_{\ln(1-W_j)}}{\sigma_{\ln(1-W_j)}}$$

$$H' = \frac{e\sum_{j=1}^{k}\psi'_j}{k}$$

$$\psi'_j = \left[\Phi^{-1}\left(1/2 \times \Phi(-z'_j)\right)\right]^2$$

$$Z'_j = \frac{\ln(1-W'_j) - \mu_{\ln(1-W'_j)}}{\sigma_{\ln(1-W'_j)}}$$

The formulas for obtaining the mean (μ) and standard error (σ) for $\ln(1 - W_j)$ are given in Royston [3] and for $\ln(1-W'_j)$ in Royston [4], and are shown in Section 2.3.

$$e = \frac{k}{1+(k-1)\overline{c}}$$

$$\overline{c} = \sum_{i=1}^{k}\sum_{j\neq i=2}^{k}\frac{c_{ij}}{k(k-1)/2} = \frac{2\sum_{i=1}^{k}\sum_{j\neq i=2}^{k}c_{ij}}{k(k-1)} = \frac{2(c_{12}+c_{13}+\cdots+c_{k-1,k})}{k(k-1)}$$

The k-order square matrix of sample correlations is required.

$$R = \begin{pmatrix} r_{11} & \cdots & r_{1j} & \cdots & r_{1k} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ r_{i1} & \cdots & r_{ij} & \cdots & r_{ik} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ r_{k1} & \cdots & r_{kj} & \cdots & r_{kk} \end{pmatrix}$$

$$c_{ij} = r_{ij}^{\lambda}\left[1 - \frac{\mu}{\upsilon_n}(1-r_{ij})^{\mu}\right]$$

If $10 \leq n \leq 2000$, then $\lambda = 5$, $\mu = 0.715$, and
$\upsilon_n = 0.21364 + 0.015124(\ln n)^2 - 0.0018034(\ln n)^3$.

The sampling distribution of the test statistic H or H' is a chi-square distribution with e degrees of freedom: $H \sim \chi_e^2$ and $H' \sim \chi_e^2$

The statistical decision under the null hypothesis of multivariate normality at a given significance level is taken one-sided to the right tail.

If $P(\chi_e^2 \geq H) \geq \alpha$, H_0 is accepted, and $< \alpha$, is rejected.

If $P(\chi_e^2 \geq H') \geq \alpha$, H_0 is accepted, and $< \alpha$, is rejected.

The size of the effect is estimated using the eta-squared coefficient.

$$\eta^2 = \frac{H}{n\times gl} = \frac{H}{n\times e} \quad \eta^2 = \frac{H'}{n\times gl} = \frac{H'}{n\times e}$$

Type-II error or probability β and statistical power ϕ are calculated using the cumulative distribution function of a non-central chi-square distribution ($NC\chi^2$): degrees of freedom (df) = e, and non-centrality parameter (NCP) = $\eta^2 \times n \times e$ = H. Its argument is the critical value for the statistics h or h: $_{1-\alpha}\chi_e^2$.

$$\beta = NC\chi_{df=e=\frac{k}{1+(k-1)\overline{c}},NCP=\eta^2\times n\times e}^2\left(_{1-\alpha}\chi_e^2\right)$$

$$\phi = 1 - \beta$$

3.2.3. Runs Test for Multivariate Normality

A *network diagram* is a set of points connected by non-directional lines. In this type of diagram, the points are called *nodes* and the lines connecting the points are called *edges*. Each edge has a weight or value corresponding to the distance between the two nodes it connects, and is denoted by w_i. In the present work, such weight is calculated through the Euclidean distance formula. Let the random vectors $\vec{x}$ and $\vec{y}$.

$$\vec{x} = (x_1 \quad x_2 \quad \cdots \quad x_k)$$

$$\vec{y} = (y_1 \quad y_2 \quad \cdots \quad y_k)$$

$$w_i = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_k - y_k)^2}$$

A *spanning tree* is a subset of nodes connected without the chain of connections returning to the origin. When the chain of connections is closed, returning to the origin, it is referred to as a *cycle*. As the nodes are numbered from 1 to n (random sampling order of the n score vectors or n data tuples), the *root* of the spanning tree is the node that receives no connections. A *minimum spanning tree* is the one that has the smallest sum of the weights of its edges. This may be unique or there may be more than one in a network diagram [8].

In the multivariate runs test applied to multivariate normality [8] [9], the random vector $\vec{x}$ corresponds to sample data and the random vector $\vec{y}$ is generated from a multivariate normal distribution, whose location parameter $\vec{\mu}$ is estimated with the vector of sample means $\vec{x}$ and its scale parameter Σ is estimated with the sample covariance matrix S of the vector $\vec{x}$.

$$Z = (R - \mu_R)/\sigma_R$$

$$\mu_R = \frac{n_0 n_1}{n} + 1$$

$$\sigma_R = \frac{2n_0 n_1}{n(n-1)}\left[\frac{2n_0 n_1 - n}{n} + \frac{c - n + 2}{(n-2)(n-3)}\left(n(n-1) - 4n_0 n_1 + 2\right)\right]$$

n_0 = the number of empirical k-dimensional tuples or k-dimensional vectors.

n_1 = the number of theoretical or generated tuples under a multivariate normal distribution model with the sample mean vector $\vec{x}$ as the estimator of the location parameter $\vec{\mu}$ and the sample covariance matrix of the vector $\vec{x}$ as the estimator of the scale parameter Σ.

$$n = n_0 + n_1 = 2 \times n_0$$

r = the number of separate trees that result when any edge (line) in the minimum spanning tree between nodes of different (empirical vs. theoretical) samples is removed.

c = the number of pairs of edges (links or lines) sharing a common node (point) in the network diagram (points joined by lines).

$$c = \frac{1}{2}\sum_{i=1}^{n} d_i(d_i - 1)$$

d_i = the degree of node i (i = 1, 2, … n) in the minimum spanning tree within the network diagram or number of edges connected to node i.

The n_1 theoretical tuples are generated from the vector of sample means, the sample covariance matrix and the lower triangular matrix of the Cholesky's decomposition of the sample covariance matrix.

Vector of sample means: $\vec{m} = (\overline{x}_1 \quad \overline{x}_2 \quad \cdots \quad \overline{x}_k) \in \mathbb{R}^k$

Sample covariance matrix:

$$S = \begin{pmatrix} s_1^2 & \cdots & s_{1k} \\ \vdots & \ddots & \vdots \\ s_{k1} & \cdots & s_k^2 \end{pmatrix} \in \mathbb{R}^{k \times k}$$

Lower triangular matrix of the Cholesky's decomposition of the sample covariance matrix:

$$C = \begin{pmatrix} c_{11} & 0 & \cdots & 0 \\ c_{21} & c_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ c_{k1} & c_{k2} & \cdots & c_{kk} \end{pmatrix} \in \mathbb{R}^{k \times k}$$

$$S = C^{\mathrm{T}}C$$

$$\vec{z} = \begin{pmatrix} z_1 & z_2 & \cdots & z_k \end{pmatrix} \in \mathbb{R}^k, \forall z_i \sim N(0,1)$$

Vector of scores generated from multivariate normal distribution:

$$\vec{y} = \vec{m} + C\vec{z} \in \mathbb{R}^k \sim N(\vec{\mu}, \Sigma)$$

The test statistic and its sampling distribution which is a standard normal distribution.

$$Z = (R - \mu_R)/\sigma_R \sim N(0,1)$$

The statistical decision under the null hypothesis of multivariate normality at a given significance level is made for a left-tailed test. If $P(Z \leq z) \geq \alpha$, H_0 is accepted; and $< \alpha$, H_0 is rejected.

The effect size can be estimated using Rosenthal's r correlation coefficient.

$$r = |z|/\sqrt{n} = |z|/\sqrt{n_1 + n_2}$$

The calculation of the type II error or probability β and the statistical power ϕ is left-tailed in accordance with the calculation of the probability value or critical level.

$$R_1 = \mu_R + z_\alpha \sigma_R$$

$$\beta = P\left(Z \geq z_1 = \frac{R_1 - R}{\sigma_R}\right) = 1 - P\left(Z \leq z_1 = \frac{R_1 - R}{\sigma_R}\right)$$

$$\phi = 1 - \beta = P(Z \leq z_1)$$

4. Results

4.1. Comparison of the Probability of the Correct Decision Conditional on the Alternative Hypothesis among the Six Multivariate Normality Tests

Table 7 shows the statistical power values ϕ when the generated samples are drawn from a population without multivariate normality, corresponding to 40 random samples. The tests are expected to reject the null hypothesis of multivariate normality. Probability β appears when the generated samples have been drawn from a population with multivariate normality (sample #6 = 4N independent, sample #7 = 4N related, sample #32 = 4N related, sample #33 = 4N related, sample #34 = 4N related, sample #41 = 4N independent, and sample #46 = 4N independent) or with a good convergence to multivariate normality (sample #38 = 4t[ν = 100] independent, sample #39 = 4χ^2[ν = 100] independent, and

Table 7. Probability of the correct decision conditional on the alternative hypothesis of the six multivariate normality tests.

	Sample	Pr.	Proposed		Mardia	SJ	Royston	
			Q	Q	K^2	Z	H	H
1	$N_1E_1C_1L_1$	ϕ	0.9996	0.9989	1	0.9998	0.9219	0.8884
2	$N_2E_2C_2L_2$	ϕ	1	0.9999	1	0.9993	0.9451	0.9244
3	$N_3E_3C_3L_3$	ϕ	1	1	1	0.9978	1	0.9999
4	$N_4E_4C_4L_4$	ϕ	1	1	0.9995	0.9999	0.9730	0.9582
5	$N_1E_2C_3L_4$	ϕ	1	1	0.9890	0.6167	1	0.9999
6	$N_1N_2N_3N_4$	β	0.7516	0.8587	0.2687	0.9755	0.6176	0.6185
7	4N rel.	β	0.9334	0.9500	0.3671	0.8404	0.8677	0.8492
8	4E ind	ϕ	1	1	0.9974	0.7330	1	0.9999
9	4C ind	ϕ	1	1	0.9849	0.8269	1	1
10	4L ind	ϕ	1	1	0.9999	0.8264	1	1
11	$N_1N_2E_1C_1$	ϕ	0.9975	0.9793	0.9998	0.9993	0.8687	0.8171
12	$N_1N_2E_1L_1$	ϕ	0.9885	0.9968	0.9971	0.9998	0.7969	0.7385
13	$N_1N_2C_1L_1$	ϕ	0.9985	0.9888	1	0.9999	0.9175	0.8889
14	$N_1N_3E_2C_2$	ϕ	0.9967	0.9621	0.9970	0.9981	0.9623	0.9490
15	$N_1N_3E_2L_2$	ϕ	0.9975	0.9976	0.9996	0.9876	0.9355	0.9154
16	$N_1N_3C_2L_2$	ϕ	0.9998	0.9942	0.9998	0.6161	0.9899	0.9828
17	$N_1N_4E_3C_3$	ϕ	1	1	1	0.9979	1	0.9998
18	$N_1N_4E_3L_3$	ϕ	1	1	0.9702	0.9946	0.9995	0.9982
19	$N_1N_4C_3L_3$	ϕ	1	1	1	0.9870	1	0.9999
20	$N_2N_3E_4C_4$	ϕ	0.9800	0.9545	0.9904	0.8979	0.9766	0.9663
21	$N_2N_3E_4L_4$	ϕ	1	1	0.9798	0.0500	0.9816	0.9898
22	$N_2N_3C_4L_4$	ϕ	0.9988	0.9978	0.9938	0.4893	0.9909	0.9846
23	$N_2N_4E_1C_3$	ϕ	1	1	0.8843	0.9872	0.9989	0.9964
24	$N_2N_4E_1L_3$	ϕ	1	1	0.9666	0.7357	0.9994	0.9976
25	$N_2N_4C_1L_3$	ϕ	1	1	0.9865	0.3634	0.9996	0.9987
26	$N_3N_4E_2C_1$	ϕ	1	0.99998	0.9321	0.3630	0.9738	0.9667
27	$N_3N_4E_2L_1$	ϕ	0.9951	0.9954	0.9393	0.6199	0.9915	0.9874
28	$N_3N_4C_2L_1$	ϕ	0.9997	0.9983	0.9953	0.3637	0.9907	0.9859
29	3NE rel	ϕ	0.5444	0.5842	0.4909	0.2515	0.5685	0.4633
30	3NC rel	ϕ	0.9572	0.8650	0.7551	0.1591	0.6984	0.6555
31	3NL rel	ϕ	0.6181	0.6983	0.6562	0.9432	0.8009	0.7332
32	4N rel	β	0.8822	0.8796	0.2782	0.6377	0.9013	0.9077

Continued

33	4N rel	β	0.9082	0.9249	0.0612	0.8409	0.8526	0.8496
34	4N rel	β	0.8344	0.6814	0.4174	0.7505	0.6934	0.7279
35	4Logist ind	ϕ	0.3846	0.3538	0.7429	0.0109	0.5109	0.5171
36	4Lap ind	ϕ	0.6833	0.7130	0.8456	0.0109	0.8250	0.8092
37	4B (10, 0.1) ind	ϕ	1.0000	1.0000	0.7956	0.0108	0.9999	0.9996
38	4t [100] ind	β	0.7194	0.6863	0.4734	0.9998	0.8090	0.7993
39	$4\chi^2$ [100] ind	β	0.6423	0.6945	0.4215	0.9890	0.7484	0.7466
40	4B (20, 0.5) ind	β	0.3549	0.3628	0.3415	0.9891	0.4726	0.5652
41	4N ind	β	0.6360	0.6750	0.2585	0.7481	0.6993	0.7342
42	$2N2N^{-1}$ ind	ϕ	1	1	1	0.9993	1	0.9999
43	$2N2\chi^2$ (2, 4) ind	ϕ	0.7991	0.8737	0.8159	0.0936	0.9203	0.8948
44	$2N2\chi^2$ [1] ind	ϕ	0.9991	0.9997	0.9166	0.1598	0.9978	0.9947
45	2N2C ind	ϕ	0.9996	0.9978	0.6199	0.2506	0.9750	0.9601
46	4N ind	β	0.4665	0.3009	0.5901	0.9755	0.8038	0.8033
47	$3N1N^{-1}$ ind	ϕ	1	1	0.9999	0.9998	0.9995	0.9981
48	$3N1\chi^2$ [2] ind	ϕ	0.7880	0.8968	0.3631	0.2504	0.8121	0.7306
49	$3N1\chi^2$ [1] ind	ϕ	1	0.9972	0.9449	0.0500	0.9877	0.9733
50	3N1C ind	ϕ	1	1	0.9872	0.9946	0.9946	0.9868

Note: Multivariate samples extracted from: N = standard normal distribution, E = exponential distribution with inverse scale parameter $\lambda = \frac{1}{2}$, C = standard Cauchy distribution, L = standard LogNormal distribution, B(n, π_{succes}) = binomial distribution with parameters n (number of trials) and p (probability of success), $\chi^2[\nu]$ = chi-square distribution with ν degrees of freedom, t[ν] = Student's t distribution with ν degrees of freedom, Lap = standard Laplace distribution, N^{-1} = inverse normal distribution, and Logist = standard logistic distribution; ind = independent variables, rel = correlated variables. The preceding number is the number of variables with the same type of distribution. When the number of the variable subscripts matches, the corresponding variables are correlated. Pr = the probability of the correct decision conditional on the alternative hypothesis; when the null hypothesis of multivariate normality must be accepted, this probability is the type II error or beta probability (β), and when the null hypothesis must be rejected, it is the complement of the beta probability or statistical power (ϕ). Tests: the proposed Q-test from the Shapiro-Wilk W-statistics [3] and the proposed Q' test from the Shapiro-Francia W' statistics [4], Mardia: $K^2 = Q + Z^2$ = multivariate normality omnibus statistic [7], SJ = multivariate runs test applied to the multivariate normality by Smith and Jain [8] [9], which is a left-tailed Z-test, and the Royston's H-test from the Shapiro-Wilk W-statistics [3] [10] and the H'-test from the Shapiro-Francia W' statistics [4] [10]. Probability values were rounded to 4 decimal places, so the 1's are an artifact of rounding.

sample #40 = 4B[n = 20, π_{success} = 0. 5] independent). With these 10 random samples, the tests are expected to sustain the null hypothesis of multivariate normality.

None of the six distributions of the β or ϕ probabilities corresponding to the six multivariate normality tests applied to the 50 generated multivariate samples followed a univariate normal distribution checked using the Shapiro-Wilk W-test (Royston, 1992) and D'Agostino-Berlanger-D'Agostino K^2-test at a 5% significance level. The distributions showed a U-shaped profile with elevations at

both ends and a depression in the center. All of them had negative skewness (left tail longer than right tail), three had positive excess kurtosis or heavy tails (Q, Q' and H'), one had negative excess kurtosis or shortened tails (Smith-Jain), and two had zero excess kurtosis (Mardia and H). See Table 8 and Table 9.

Nor did any of the ϕ probabilities of the six multivariate normality tests applied to the 40 multivariate samples drawn from populations without multivariate normality followed a univariate normal distribution either by the Shapiro-Wilk W-test [3] or D'Agostino-Berlanger-D'Agostino K^2-test at the 5% significance level [47]. Five of their distributions showed negative skewness with an elongated profile of increasing staircase slope to the right, except for the distribution of ϕ probabilities from the runs test which showed symmetry. The profile of five of the six distributions was leptokurtic or thick-tailed, except for the profile of the runs test, which was platykurtic or thin-tailed. See Table 8 and Table 9.

However, the distributions of the β probabilities of four of the six multivariate normality tests applied to the 10 multivariate samples (drawn from populations with multivariate normality or with good convergence to it) were fitted to a

Table 8. Check of univariate normality using the Shapiro-Wilk test.

n	Statistic	Proposal		Mardia's	SJ	Royston's	
		Q	Q	K^2	Z	H	H'
50	W	0.6494	0.6325	0.7488	0.8046	0.7675	0.8001
	Z	5.9769	6.0772	5.2658	4.7306	5.1013	4.7790
	p	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	Normality	no	no	no	no	no	no
40	W	0.4679	0.4751	0.6446	0.8279	0.6718	0.7056
	Z	6.2316	6.1915	5.5607	4.0350	5.3934	5.1648
	p	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	Normality	no	no	no	no	no	no
10	W	0.9287	0.8735	0.8765	0.7942	0.9634	0.9746
	Z	0.1642	1.2270	1.1811	2.2472	−0.9280	−1.4755
	p	0.4348	0.1099	0.1188	0.0123	0.8233	0.9300
	Normality	yes	yes	yes	no	yes	yes

Note. n = sample size, W = the Shapiro-Wilk test statistic, Z = the standardized value of log-transformed W statistics using the Royston's formulas [3], p = right-tailed probability value under standard normal distribution, and Normality: yes when $p < \alpha = 0.05$ for n = 50 and 40 and 0.1 for n = 10, no when $p \geq \alpha$. Multivariate normality tests: the proposed Q-test from the Shapiro-Wilk W-statistics [3] and the Q'-test from the Shapiro-Francia W' statistics [4], Mardia: $K^2 = Q + Z^2$ = multivariate normality omnibus statistic [7], SJ = multivariate runs test applied to the multivariate normality by Smith and Jain [8] [9], which is a left-tailed Z-test, and the Royston's H-test from the Shapiro-Wilk W-statistics [3] [10] and the H'-test from the Shapiro-Francia W' statistics [4] [10].

Table 9. Tests of univariate skewness, kurtosis and normality based on standardized central moments.

n	Stat.	Proposal		Mardia's	SJ	Royston's	
		Q	Q'	K^2	Z	H	H'
50	$z(\sqrt{b_1})$	−4.1668	−4.6032	−3.1285	−2.2595	−3.7671	−3.4225
	p	<0.0001	<0.0001	0.0018	0.0239	0.0002	0.0006
	$z(b_2)$	2.2434	2.9631	0.2025	−2.1260	1.8991	1.4236
	p	0.0249	0.0030	0.8395	0.0335	0.0576	0.1546
	K^2	22.3948	29.9696	9.8284	9.6251	17.7975	13.7400
	p	<0.0001	<0.0001	0.0073	0.0081	0.0001	0.0010
	$N(\mu, \sigma)$	no: − & ↑	no: − & ↑	no: −	no: − & ↓	no: − & ↑	no: −
40	$z(\sqrt{b_1})$	−5.0475	−5.4837	−4.4467	−1.3323	−4.2865	−4.0234
	p	0.0000	0.0000	<0.0001	0.1828	<0.0001	0.0001
	$z(b_2)$	3.7356	4.3276	3.1058	−4.0867	2.9193	2.5986
	p	0.0002	0.0000	0.0019	<0.0001	0.0035	0.0094
	K^2	39.4326	48.7984	29.4190	18.4761	26.8963	22.9408
	p	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	$N(\mu, \sigma)$	no: − & ↑	no: − & ↑	no: − & ↑	no: ↓	no: − & ↑	no: − & ↑
10	$z(\sqrt{b_1})$	−1.0754	−1.2556	1.9175	−1.7229	−1.0924	−0.3804
	p	0.2822	0.2093	0.0552	0.0849	0.2747	0.7037
	$z(b_2)$	−0.0624	0.0740	1.9760	0.4832	0.6720	0.1697
	p	0.9503	0.9410	0.0481	0.6290	0.5016	0.8653
	K^2	1.1605	1.5820	7.5815	3.2018	1.6449	0.1735
	p	0.5598	0.4534	0.0226	0.2017	0.4394	0.9169
	$N(\mu, \sigma)$	yes	yes	no: + & ↑	yes	yes	yes

Note. n = sample size, $z(\sqrt{b_1})$ = standardized value of skewness measure based on standardized third central moment [50], p = two-tailed probability under a standard normal distribution, z (b_2) = standardized kurtosis measure value based on standardized fourth central moment [51], p = two-tailed probability value under a standard normal distribution, $K^2 = z(\sqrt{b_1})^2 + z(b_2)^2$ = test statistic from the D'Agostino-Belanger-D'Agostino omnibus test of normality [47], p = probability to the right tail under a chi-square distribution with two degrees of freedom. $N(\mu, \sigma)$ = fit to normality: yes when $p < \alpha = 0.05$ for n = 50 and 40 and 0.1 for n = 10, no when $p < 0.05$, either due to skewness (+positive or −negative) or kurtosis (↓ b2 < 3 leptokurtosis or ↑ b_2 > 3 platykurtosis).

model a univariate normality distribution checked using W- and K^2-tests at the 10% significance level (Q, Q', H and H'). The distribution of β probabilities from the runs test deviated from normality when tested using the W-test and showed

negative skewness. The distribution of β probabilities from the Mardia's test did not follow a normal distribution according to the K^2-test, as they showed a positive skew and a leptokurtic profile. See Table 8 and Table 9. Except for the two versions of the Royston test, no profile of β probabilities was bell-shaped in the histogram. The profiles were ladder-shaped with their steps increasing with Q-, Q'-, and runs z-test and decreasing with Mardia's test. Nor does the dotted line clearly lined up at 45 on the normal quantile-quantile plot with these last four tests.

When comparing probabilities β or ϕ, the choice was made to use nonparametric tests with the total of 50 generated multivariate samples and the 40 samples drawn from populations without multivariate normality. Even a non-parametric analysis approach was adopted with the 10 samples drawn from populations with multivariate normality or a good convergence to it due to the small sample size [48], consistency with previous analyses, and lack of evidence of normality in the plots [49], particularly in four of them.

Using the Friedman's omnibus test for comparing values ϕ or β, there was a significant difference both in the sample of 50 tuples and in the sample of 40 tuples from a population without multivariate normality and in the sample of 10 tuples from a population with multivariate normality or a good convergence to it ($Q > 11.071$, $p < 0.05$). The statistical power was very high in the three tests ($\phi > 0.9$). Following the cut-off points suggested by Kendall and Gibbons [32], the effect size, estimated by the Kendall's coefficient of agreement, was small in the 50-tuple and 10-tuple samples, with values in the interval (0.1, 0.3), and reached a medium level in the sample of 40 tuples (See Table 10).

In the random sample of 50 tuples, the highest median probability of the correct decision conditional on the alternative hypothesis (ϕ or β values) appeared with the version from the Shapiro-Francia statistics of the proposed test, Mdn (ϕ or β) = 0.9989, followed by the same test from the Shapiro-Wilk W'-statistics, Mdn (ϕ or β) = 0.9974. Thirdly, the Royston's H-test from the Shapiro-Wilk W statistics was located, Mdn (ϕ or β) = 0.9734, fourthly, the Mardia's K^2-test, Mdn (ϕ or β) = 0.9684, fifthly, the Royston's H-test from the Shapiro-Francia W' statistic, Mdn (ϕ or β) = 0.9591, and lastly the runs Z-test, Mdn (ϕ or β) = 0.8337.

When making the pairwise comparisons through the Wilcoxon's signed-rank test from the asymptotic probability with the Sidak's correction using Holm's procedure to control for the family rate error [35] [36], the two versions of the proposed test were equivalent in ϕ or β values and both were superior to the runs Z-test, Mardia's K^2-test, and Royston H'-test from Shapiro-Francia W' statistics. The version of the Royston test from the Shapiro-Wilk W-statistics was superior to the version from the Shapiro-Francia W'-statistics and was also superior to the runs Z-test and Mardia's K^2-test. In turn, the Royston H'-test from Shapiro and Francia W' statistic was superior to the runs Z-test. Following the cut-off points suggested by Cohen (1988), the effect sizes estimated using Rosenthal's r coefficient varied from small (0.1 to 0.29) to medium (0.3 to 0.49). See Table 11.

Table 10. Friedman test, effect size, and statistical power.

Variable	n	Q	df	p	Kendall's W	$_{0.95}\chi^2_{(5)}$	β	ϕ
ϕ o β	50	52.7123	5	3.85×10^{-10}	0.2116	11.0705	7.34×10^{-6}	0.99999
ϕ	40	73.2948	5	2.11×10^{-14}	0.3665	11.0705	1.08E-08	0.91792
β	10	13.8286	5	0.0167	0.2766	11.0705	0.16664	0.90243

Note: Compared variable: ϕ or β = the probability of the correct decision conditional on the alternative hypothesis of the six multivariate normality tests (ϕ with the 40 samples drawn from distributions without multivariate normality and β with the 10 samples from distributions with multivariate normality or a good convergency to it), ϕ = statistical power or probability of rejecting the null hypothesis when the alternative hypothesis is true, β = type II error or probability of maintaining the null hypothesis when the alternative hypothesis is true, n = the number of tuples, Q = the value of the Friedman's test statistic, df = the degrees of freedom, p = right-tailed probability in a chi-square distribution with five degrees of freedom, Kendall's W = the coefficient of agreement Kendall's W as a measure of effect size, $_{0.95}\chi^2_{(5)}$ (5) = critical value or quantile of order 0.95 of a chi-square distribution with five degrees of freedom, β = type II error, and Φ = statistical power.

Table 11. Pairwise comparisons between the six multivariate normality tests using the Wilcoxon signed-rank test in the sample of 50 tuples.

G_1	G_2	Mdn_{G1}	Mnd_{G2}	up	z	p	r	i	α_c	Sig.	Dif.
Q'	Q	0.999	0.997	38	<0.001	1	0	15	0.0500	No	
Q'	M	0.999	0.968	49	−3.482	0.0005	0.348	4	0.0136	Sí	Q' > M
Q'	SJ	0.999	0.834	50	−3.774	0.0002	0.377	2	0.0068	Sí	Q' > SJ
Q'	H	0.999	0.973	50	−1.786	0.0741	0.179	14	0.0467	No	
Q'	H'	0.999	0.959	50	−2.365	0.0180	0.237	9	0.0303	Sí	Q' > H'
Q	M	0.997	0.968	49	−3.561	0.0004	0.356	3	0.0102	Sí	Q > M
Q	SJ	0.997	0.834	50	−3.456	0.0005	0.346	5	0.0170	Sí	Q > SJ
Q	H	0.997	0.973	50	−1.824	0.0681	0.182	13	0.0435	No	
Q	H'	0.997	0.959	50	−2.182	0.0291	0.218	10	0.0336	Sí	Q > H'
M	SJ	0.968	0.834	50	−1.902	0.0572	0.190	12	0.0402	No	
M	H	0.968	0.973	50	−2.597	0.0094	0.260	8	0.0270	Sí	H > M
M	H'	0.968	0.959	50	−2.018	0.0436	0.202	11	0.0369	No	
SJ	H	0.834	0.959	50	−2.973	0.0029	0.297	6	0.0203	Sí	H > SJ
SJ	H'	0.834	0.959	50	−2.838	0.0045	0.284	7	0.0237	Sí	H' > SJ
H	H'	0.973	0.959	50	−4.035	0.0001	0.404	1	0.0034	Sí	H > H'

Note: G_1 = group 1 and G_2 = group 2. Multivariate normality tests: the Q = proposed Q-test from the Shapiro-Wilk W statistics and the Q'-test from the Shapiro-Francia W' statistics, M = the Mardia's K^2-test, SJ = the Smith-Jain runs Z-test, H = the Royston H-test from the Shapiro-Wilk W statistics, and the H'-test from the Shapiro-Francia W' statistics, *Mdn* (ϕ or β) = the sample median of ϕ or β values, up = the number of unequal pairs (non-zero differences), z = the Wilcoxon's rank-signed test z-statistic from its approximation to the normal distribution with the correction for ties and the continuity correction, p = two-tailed probability under a standard normal distribution, $r = |z|/\sqrt{100}$ = the Rosenthal's r coefficient as a measure of effect size, i = range in ascending order of probability value with average ranges in case of ties, $\alpha_c = 1 - 0.95^{i/15}$ = significance level with the Sidak's correction using the Holm's procedure [35] [36], Sig = significance: no when $p \geq \alpha_c$ and yes when $p < \alpha_c$, Diff = the group with the highest median in each pairwise comparison.

In the random sample of 40 tuples, the highest median statistical power (ϕ) appeared with the version from the Shapiro-Francia W' statistics of the proposed test, $Mdn(\phi) = 0.9999$, followed by its version from Shapiro-Wilk W-statistics, $Mdn(\phi) = 0.9993$. Thirdly, the median of the Mardia's K^2-test was located, $Mdn(\phi) = 0.9868$, fourthly, that of the Royston's H-test from the Shapiro-Wilk W statistics, $Mdn(\phi) = 0.9847$), fifthly, that of the Royston's H'-test from the Shapiro-Francia W' statistics, $Mdn(\phi) = 0.9781$, and lastly that of runs Z-test, $Mdn(\phi) = 0.7419$. When making the pairwise comparisons using the Wilcoxon signed-rank test from the asymptotic probability with the Sidak's correction with the Holm's procedure [35] [36], the two versions of the proposed test were equivalent in ϕ values and both were superior to the other tests. Again, the version of the Royston's test from the Shapiro-Wilk-statistics was superior to the version from the Shapiro-Francia W' statistics with a large effect size (Rosenthal's $r = 0.506$) and both versions were superior to the runs Z-test. In turn, the Mardia's K^2-test was superior to the runs Z-test. The effect sizes estimated using Rosenthal's r coefficient varied from small (from 0.1 to 0.29) to medium (from 0.3 to 0.49), reaching a large effect size on the statistical power when choosing between the proposed Q'-test and the runs Z-test. See Table 12.

In the random sample of 10 tuples, the highest median type II error (β) appeared with the runs Z-test, $Mdn(\beta) = 0.9755$. Secondly, the Royston's H-test from the Shapiro-Wilk W-statistics was located, $Mdn(\beta) = 0.8064$, thirdly, this same test from the Shapiro-Francia W'-statistics, $Mdn(\beta) = 0.8013$, fourthly, the proposed test from the Shapiro-Francia W' statistics, $Mdn(\beta) = 0.7355$, fifthly, this same test from the Shapiro-Wilk W-statistics, $Mdn(\beta) = 0.6904$, and lastly the Mardia's K^2-test, $Mdn(\beta) = 0.3922$. When pairwise comparisons were made using the Wilcoxon signed-rank test from the exact probability with Sidak's correction with Holm's procedure with a nominal significance level of 0.1, the two versions of the proposed test were equivalent in β values as were the two versions of Royston's test. The Mardia's K^2-test had a significantly lower β probability values than the two versions of the Royston test and the runs Z-test. The effect sizes estimated using the rank biserial correlation were very large. See Table 13.

4.2. Difference in the Number of Successes among the Six Multivariate Normality Tests

The proposed Q-test from the Shapiro-Wilk W statistics and the Royston's H-test also based on the Shapiro-Wilk W statistics had the highest proportion of correct classifications or probability of success ($p_s = 0.94$). Secondly, Royston's H'-test from the Shapiro-Francia W'-statistics was located ($p_s = 0.94$), thirdly, the proposed Q'-test from the Shapiro-Francia W' statistics ($p_s = 0.88$), fourthly, the Mardia's K^2-test ($p_s = 0.70$), and lastly, the runs Z-test ($p_s = 0.68$).

Using Cochran's omnibus Q test, there was significant difference in the sample of 50 tuples ($Q = 44.593$, $df = 5$, $p < 0.001$). The statistical power of the test was very high ($\phi = 0.9999$). Following the cut-off points suggested by Cohen

Table 12. Pairwise comparisons between the six multivariate normality tests using the Wilcoxon signed-rank test on the sample of 40 tuples.

G_1	G_2	Mdn_{G1}	Mdn_{G2}	up	z	p-value	r	i	α_c	Sig.	Diff.
Q'	Q	0.9999	0.9993	28	−0.512	0.60840	0.057	14	0.0467	No	
Q'	M	0.9999	0.9868	39	−2.561	0.01044	0.286	12	0.0369	Sí	Q' > M
Q'	SJ	0.9999	0.7419	40	−4.751	<0.00001	0.531	1	0.0051	Sí	Q' > SJ
Q'	H	0.9999	0.9847	40	−2.991	0.00278	0.334	10	0.0402	Sí	Q' > H
Q'	H'	0.9999	0.9781	40	−3.676	0.00024	0.411	7	0.0270	Sí	Q' > H'
Q	M	0.9993	0.9868	39	−2.728	0.00637	0.305	11	0.0336	Sí	Q > M
Q	SJ	0.9993	0.7419	40	−4.563	0.00001	0.510	2	0.0102	Sí	Q > SJ
Q	H	0.9993	0.9847	40	−3.246	0.00117	0.363	9	0.0303	Sí	Q > H
Q	H'	0.9993	0.9781	40	−3.703	0.00021	0.414	6	0.0237	Sí	Q > H'
M	SJ	0.9868	0.7419	40	−3.985	0.00007	0.446	4	0.0136	Sí	M > SJ
M	H	0.9868	0.9847	40	−0.880	0.37864	0.098	13	0.0435	No	
M	H'	0.9868	0.9781	40	−0.114	0.90904	0.013	15	0.0500	No	
SJ	H	0.7419	0.9847	40	−3.717	0.00020	0.416	5	0.0170	Sí	H > SJ
SJ	H'	0.7419	0.9781	40	−3.555	0.00038	0.397	8	0.0203	Sí	H' > SJ
H	H'	0.9877	0.9781	40	−4.523	0.00001	0.506	3	0.0051	Sí	H' > H

Note: G_1 = group 1 and G_2 = group 2. Multivariate normality tests: Q = the proposed Q-test from the Shapiro-Wilk W statistics and the Q'-test from the Shapiro-Francia W' statistics, M = the Mardia's K^2-test, SJ = the Smith-Jain runs Z-test, H = the Royston H-test from the Shapiro-Wilk W statistics, and the H'-test from the Shapiro-Francia W' statistics, Mdn (ϕ) = the sample median of ϕ values, up = the number of unequal pairs (non-zero differences), z = the Wilcoxon's rank-signed test z-statistic from its approximation to the normal distribution with the correction for ties and the continuity correction, p = two-tailed probability under a standard normal distribution, $r = |z|/\sqrt{80}$ = the Rosenthal's r coefficient as a measure of effect size, i = range in ascending order of probability value with average ranges in case of ties, $\alpha_c = 1 - 0.95^{i/15}$ = significance level with the Sidak's correction using the Holm's procedure [35] [36], Sig = significance: no when $p \geq \alpha_c$ and yes when $p < \alpha_c$, Diff = the group with the highest median in each pairwise comparison.

Table 13. Pairwise comparisons between the six multivariate normality tests using the Wilcoxon signed-rank test on the sample of 10 tuples.

G_1-G_2	Mdn_1	Mdn_2	up	SR−	SR+	T	p	r_{bp}	i	α_c	Sig.	>
Q'-Q	0.736	0.690	10	30	25	25	0.8457	0.091	14	0.0937	No	
Q'-M	0.736	0.392	10	7	48	7	0.0371	0.745	4	0.0277	No	
Q'-SJ	0.736	0.975	10	44	11	11	0.1055	0.600	6.5	0.0446	No	
Q'-H	0.736	0.806	10	35	20	20	0.4922	0.273	13	0.0873	No	
Q'-H'	0.736	0.801	10	36	19	19	0.4316	0.309	12	0.0808	No	
Q-M	0.690	0.392	10	11	44	11	0.1055	0.600	6.5	0.0446	No	
Q-SJ	0.690	0.975	10	45	10	10	0.0840	0.636	5	0.0345	No	

Continued

Q-H	0.690	0.806	10	38	17	17	0.3223	0.382	10.5	0.0711	No	
Q-H'	0.690	0.801	10	38	17	17	0.3223	0.382	10.5	0.0711	No	
M-SJ	0.392	0.975	10	54	1	1	0.0039	0.964	2	0.0139	Yes	SJ > M
M-H	0.392	0.806	10	45	0	0	0.0039	1	2	0.0139	Yes	H > M
M-H'	0.392	0.801	10	54	1	1	0.0039	0.964	2	0.0139	Yes	H' > M
SJ-H	0.975	0.806	10	14	41	14	0.1934	0.491	8.5	0.0580	No	
SJ-H'	0.975	0.801	10	14	41	14	0.1934	0.491	8.5	0.0580	No	
H-H'	0.806	0.801	10	28	27	27	1	0.018	15	0.1000	No	

Note: G_1 = group 1 and G_2 = group 2. Multivariate normality tests: Q = the proposed Q-test from the Shapiro-Wilk W statistics and the Q'-test from the Shapiro-Francia W' statistics, M = the Mardia's K^2-test, SJ = the Smith-Jain runs Z-test, H = the Royston H-test from the Shapiro-Wilk W statistics, and the H'-test from the Shapiro-Francia W' statistics, *Mdn* (β) = the sample median of β values. Wilcoxon's signed-rank test: *up* = the number of unequal pairs (non-zero differences), *SR*– = the sum of negative ranks ($G_1 < G_2$), *SR* + = the sum of positive ranks ($G_1 > G_2$), T = test statistics or minor of *SR* + and *SR*– statistics, *p* = two-tailed exact probability, r_{bp} = ($SR+ - SR-$)/($SR+ + SR-$) = biserial rank correlation as a measure of effect size, *i* = range in ascending order of probability value with average ranges in case of ties, α_c = $1 - 0.95^{i/15}$ = significance level with the Sidak's correction using the Holm's procedure [35] [36], Sig = significance: no when $p \geq \alpha_c$ and yes when $p < \alpha_c$, Diff = the group with the highest median in each pairwise comparison.

[30], the effect size estimated using the eta-squared coefficient was small (η^2 = 0.032), with a value in the interval [0.01, 0.6].

When pairwise comparisons were made using the McNemar's test from the exact probability (binomial distribution) applying the Sidak's correction using the Holm's procedure to control for the family error rate, there was no significant difference in successes between the two versions of the proposed test and the two versions of the Royston's test, these four tests being statistically equivalent in number of successes. Both versions of the proposed test and the Royston test were more correct than the Mardia's K^2-test and the runs Z-test, the latter two being equivalent to each other. When the effect size was estimated by the odds ratio (*OR*) or Cohen's g statistic and, following the cutoff points suggested by the author, the effect size was large in 6 of 8 significant differences (*OR* > 4.25 and $|g|$ > 0.25). In the other two comparisons, the *OR* remained undefined and, therefore, so did the *g*-statistic. When the effect size in these two comparisons was calculated using the eta-squared coefficient, it was also large: $\eta^2 = Q/(n_{01} + n_{10}) > 0.25$, where $Q = (|n_{01} + n_{10}| - 1)^2/(n_{01} + n_{10})$. See Table 14.

Although the difference between the two versions of the proposed test was not significant (exact probability: point value and one-tailed value of 0.125 and two-tailed value of 0.25 > α = 0.05), the effect size was large (*OR* = 0, g = −0.5, η^2 = 0.44), the type II error was null and the unit statistical power, which supports the alternative hypothesis of difference, where Q-test would have a higher proportion of successes than Q'-test (0.94 versus 0.88). Consequently, there is a situation of ambiguity with respect to this difference.

Table 14. Pairwise comparisons of the number of successes using the McNemar's test.

Test		Joint probabilities				Two-tailed exact probability				Effect size			Holm-Sidak correction			
T_1	T_2	n_{00}	n_{01}	n_{10}	n_{11}	x	n	p	ϕ	OR	g	η^2	i	α_c	Sig	$>$
Q'	Q	3	3	0	44	0	3	0.25	1	0	−0.5	0.44	9	0.0303	No	
Q'	M	6	0	9	35	0	9	0.004	I	I		0.79	7	0.0237	Yes	Q'
Q'	SJ	5	1	11	33	1	12	0.006	0.928	11	0.417	0.56	8	0.0270	Yes	Q'
Q'	H	2	4	1	43	1	5	0.375	0.328	0.25	−0.3	0.16	10	0.0336	No	
Q'	H'	2	4	2	42	2	6	0.688	0.089	0.5	−0.167	0.03	11	0.0369	No	
Q	M	3	0	12	35	0	12	0.000	I	I		0.84	3	0.0102	Yes	Q
Q	SJ	2	1	14	33	1	15	0.001	0.985	14	0.433	0.64	5.5	0.0186	Yes	Q
Q	H	2	1	1	46	1	2	1	0.5	1	0	0.25	13.5	0.0451	No	
Q	H'	2	1	2	45	1	3	1	0.333	2	0.167	0.00	13.5	0.0451	No	
M	SJ	12	3	4	31	3	7	1	0.013	1.33	0.071	0.00	13.5	0.0451	No	
M	H	3	12	0	35	0	12	0.0005	1	0	−0.5	0.84	3	0.0102	Yes	H
M	H'	4	11	0	35	0	11	0.0010	1	0	−0.5	0.83	5.5	0.0186	Yes	H'
SJ	H	3	13	0	34	0	13	0.0002	1	0	−0.5	0.85	1	0.0034	Yes	H
SJ	H'	4	12	0	34	0	12	0.0005	1	0	−0.5	0.84	3	0.0102	Yes	H'
H	H'	3	0	1	46	0	1	1	I	I		0	13.5	0.0451	No	

Note: T_1 = first test, T_2 = second test, n_{00} = the number of concordant pairs of non-normality for both tests, n_{01} = the number of discordant pairs of non-normality for the first test and normality for the second test, n_{10} = the number of discordant pairs of normality for the first test and non-normality for the second test, n_{11} = number of concordant pair of normality for both tests, x = the smaller of the discordant frequencies (n_{01} or n_{10}) or number of successes for the calculation of the exact probability, n = sum of the discordant frequencies or parameter of the number of trials for the calculation of the exact probability, p = two-tailed exact probability under a binomial distribution B($n = n_{01} + n_{10}$, p = 0.5), ϕ = two-tailed statistical power (I = undefined), $OR = n_{10}/n_{01}$ = Cohen's odds ratio for correlated 2 × 2 tables (I = undefined), Cohen's $g = n_{10}/(n_{01} + n_{10}) - 0.5$ = effect size statistic that is only calculated if the OR value is defined, $\eta^2 = Q/(n_{01} + n_{10})$ = eta-squared coefficient, calculated with the McNemar's test statistic Q with its continuity correction, i = rank in ascending order of the two-tailed exact probability values with average ranks in case of ties, α_c = significance level with the Sidak's correction using the Holm's (1979) procedure, Sig = significance: yes when $p < \alpha_c$ and no when $p \geq \alpha_c$, ≥group with the highest number of successes in the pairwise comparison.

4.3. Sensitivity, Specificity and Efficiency of the Six Multivariate Normality Tests

The six tests presented a sensitivity of 100%, so the confidence intervals were calculated using the rule of three [52]: [1 − ln(0.05)/50, 1], as can be seen in Table 15. The specificity or ability to successfully reject the null hypothesis in case of deviation from multivariate normality varied from a minimum of 0.825 with the runs Z-tests to a maximum of 0.925 with the proposed Q-test from the Shapiro-Wilk W statistics. When making interval estimates at the 95% confidence level, the Wilson's score intervals with the continuity correction [40] [41] for the specificity values of the six tests overlapped (Table 15). When making comparisons using the McNemar's Z-test for two paired samples, the null hypothesis of

Table 15. Point estimates and 95% confidence intervals for the sensitivity, specificity and efficiency of the six multivariate normality tests.

MN Test	Table 2 × 2				Sensibility			Specificity			Efficiency		
	n_{00}	n_{01}	n_{10}	n_{11}	S	LI	LS	E	LB	UB	Ef	LB	UB
Q'	34	0	6	10	1	0.940	1	0.850	0.715	0.930	0.88	0.750	0.950
Q	37	0	3	10	1	0.940	1	0.925	0.805	0.977	0.94	0.825	0.984
M	33	0	7	10	1	0.940	1	0.825	0.686	0.913	0.86	0.726	0.937
SJ	33	0	7	10	1	0.940	1	0.825	0.686	0.913	0.86	0.726	0.937
H	35	0	5	10	1	0.940	1	0.875	0.744	0.947	0.90	0.774	0.963
H'	36	0	4	10	1	0.940	1	0.900	0.774	0.963	0.92	0.799	0.974

Note: MN Test = Multivariate normality tests: Q = the proposed Q-test from the Shapiro-Wilk W statistics and the Q' test from the Shapiro-Francia W' statistics, M = the Mardia's K^2-test, SJ = the Smith-Jain runs Z-test, H = the Royston H-test from the Shapiro-Wilk W statistics, and the H' test from the Shapiro-Francia W' statistics. Joint frequencies: n_{00} = the frequency of successes when classifying as the sample as coming from the population without multivariate normal distribution, n_{01} = the frequency of false negatives, n_{10} = the frequency of false positives, and n_{11} = the frequency of successes when classifying the sample as coming from the population with multivariate normal distribution. $S = n_{11}/(n_{11} + n_{01})$ = sensitivity or proportion of successes when detecting cases of multivariate normality, $E = n_{00}/(n_{00} + n_{10})$ = specificity or proportion of successes when detecting multivariate non-normality cases, and $Ef. = (n_{00} + n_{11})/(n_{11} + n_{01} + n_{10} + n_{11})$ = efficiency or proportion of successes when classifying. LB = the lower bound and UB = the upper bound of the Wilson's score interval with Newcombe's continuity correction at 95% confidence level.

no difference was maintained at the 5% significance level, even without considering any correction for family error rate in all 15 comparisons. The mean and median of specificity values were high, 0.867 and 0.863, respectively.

The efficiency or successes ratio varied from a minimum of 0.86 with the runs Z-test to a maximum of 0.94 with the proposed Q' test from the Shapiro-Wilk W statistics. The confidence intervals of the six tests overlapped. When comparisons were made using the Z-test for two paired samples, the null hypothesis of equivalence was maintained at the 5% significance level for the 15 differences, even without considering any correction for family error rate. The mean and median of efficiency values were high, 0.893 and 0.890, respectively.

As the lowest values of specificity and efficiency were observed in the Mardia's K^2-test and runs Z-test, but did not reveal to be significant in the comparisons of proportions with the other four tests, it was chosen to test in each of the six tests whether the specificity and efficiency values are equal to or greater than 0.90. At a significance level of 5%, the null hypothesis was rejected with the Mardia's K^2-test and runs Z-test with respect to specificity value with low statistical power ($0.5 < \phi = 0.54 < 0.80$) and small effect size ($0.10 < r = 0.25 < 0.30$). Applying the Sidak's correction with the Holm's procedure, the null hypothesis would not be rejected: p (the probability value for a left-tailed hypothesis) = 0.039 ($i = 1.5$) > $\alpha_c = 1 - 0.95^{(1.5/6)} = 0.013$. In all other cases, the null hypothesis was hold (Table 16).

Table 16. Test of a value equal or greater than 0.85 for specificity and efficiency using the McNemar Z-test.

MN Test	E	H₀: E ≥ 0.9				Ef	H₀: Ef ≥ 0.9			
		z	p	φ	r		z	p	φ	r
Q'	0.850	−1.179	0.119	0.320	0.167	0.88	−0.471	0.319	0.120	0.067
Q	0.925	0.589	0.722	0.013	0.083	0.94	0.943	0.827	0.005	0.133
M	0.825	−1.768	0.039	0.549	0.250	0.86	−0.943	0.173	0.241	0.133
SJ	0.825	−1.768	0.039	0.549	0.250	0.86	−0.943	0.173	0.241	0.133
H	0.875	−0.589	0.278	0.146	0.083	0.9	0.000	0.500	0.050	0
H'	0.9	0	0.5	0.050	0.000	0.92	0.471	0.681	0.017	0.067

Note: MN test = multivariate normality test: Q = the proposed Q-test from the Shapiro-Wilk W-statistics and the Q' test from the Shapiro-Francia W' statistics, M = Mardia's K^2-test, SJ = the Smith-Jain runs Z-test, H = the Royston's H-test from the Shapiro-Wilk W statistics and the H' test from the Shapiro-Francia W' statistics. E = specificity value, *Ef* = efficiency value, z = Z-test statistic for a population proportion, p = left-tailed probability in a standard normal distribution under the null hypothesis $E \geq 0.9$ in the fourth column and $Ef \geq 0.9$ in the seventh column, ϕ = left-tailed statistical power, $r = |z|/\sqrt{50}$ = effect size measure based on the Rosenthal's r coefficient.

4.4. Correlation between the Critical Level or Probability Value and the Deviation from Normality

Finally, the deviation from multivariate normality of the 50 generated multivariate samples was classified by ordered categories, as can be seen in Table 17. The higher the level in variable D, the greater the deviation from multivariate normality.

All six correlations were significant and negative. The negative sign of the correlation means that the smaller the critical level or probability value of the test, the greater the deviation from normality according to expectation. The highest absolute correlation of the ordinal variable D (of deviation from multivariate normality) was with the probability value of the proposed Q-test from the Shapiro-Wilk W statistics ($rho_{QD} = -0.746$, 95% *CI* [−0.789, −0.399]), followed by the same test from the Shapiro-Francia W' statistics ($rho_{Q'D} = -0.740$, 95% *CI* [−0.787, −0.395]). In third place was the correlation with the Royston's H' test from the Shapiro-Francia W' statistics ($rho_{H'D} = -0.677$, 95% *CI* [−0.759, −0.345]), in fourth place, with the same test from the Shapiro-Wilk W statistics ($rho_{HD} = -0.664$, 95% *CI* [−0.753, −0.335]), and in fifth place, with the Mardia's K^2-test ($rho_{MD} = -0.645$, 95% *CI* [−0.744, −0.319]). These five correlations were statistically equivalent using the Steiger's Z-test and, following the cut-off points suggested by Cohen (1988), their strength of association was high ($r_s > 0.50$). In sixth place was the correlation with the Smith-Jain runs test ($rho_{SJD} = -0.455$, 95% *CI* [−0.638, −0.153]), with a medium strength of association. The latter presented significant differences with the two versions of the proposed test and

the Mardia's K^2-test at a 5% significance level. If the Sidak's correction using the Holm's procedure is considered, only with the two versions of the proposed test would be significant, since their strength of association with D was very high, *rho* > 0.70 (Table 18 and Table 19).

Table 17. Classification of the 50 generated multivariate samples in five ordered categories of deviation from multivariate normality.

	Level 1		Level 3		Level 4		Level 5
6	$N_1N_2N_3N_4$	29	3NE rel	11	$N_1N_2E_1C_1$	1	$N_1E_1C_1L_1$
7	4N rel	30	3NC rel	12	$N_1N_2E_1L_1$	2	$N_2E_2C_2L_2$
32	4N rel	31	3NL rel	13	$N_1N_2C_1L_1$	3	$N_3E_3C_3L_3$
33	4N rel	35	4Logist ind.	14	$N_1N_3E_2C_2$	4	$N_4E_4C_4L_4$
34	4N rel	48	$3N1\chi^2_{(2)}$ ind	15	$N_1N_3E_2L_2$	5	$N_1E_2C_3L_4$
41	4N ind	49	$3N1\chi^2_{(1)}$ ind	16	$N_1N_3C_2L_2$	8	4E ind
46	4N ind	50	3N1C ind	17	$N_1N_4E_3C_3$	9	4C ind.
	Level 2			18	$N_1N_4E_3L_3$	10	4L ind.
38	$4t_{(100)}$ ind			19	$N_1N_4C_3L_3$	36	4Lap ind.
39	$4\chi^2_{(100)}$ ind			20	$N_2N_3E_4C_4$	37	4B (10,0.1)ind
40	4B (20, 0.5) ind			21	$N_2N_3E_4L_4$	42	$2N2N^{-1}$ ind
				22	$N_2N_3C_4L_4$	47	$3N1N^{-1}$ ind
				23	$N_2N_4E_1C_3$		
				24	$N_2N_4E_1L_3$		
				25	$N_2N_4C_1L_3$		
				26	$N_3N_4E_2C_1$		
				27	$N_3N_4E_2L_1$		
				28	$N_3N_4C_2L_1$		
				43	$2N_2\chi^2_{(2, 4)}$ ind		
				44	$2N_2\chi^2_{(1)}$ ind		
				45	$2N_2C$ ind		

Note: N = standard normal distribution, E = exponential distribution with inverse scale parameter λ = ½, C = standard Cauchy distribution, L = standard LogNormal distribution, B (n, p) = binomial distribution with parameters n (number of trials) and p (probability of success), $\chi^2(\nu)$ = chi-square distribution with ν degrees of freedom, Lap = standard Laplace distribution, N^{-1} = inverse normal distribution, Logist = standard logistic distribution, ind = independent variables, rel = correlated variables. The preceding number is the number of variables of the same type of distribution. When the number of subscripts of variables matches, the corresponding variables are correlated.

Table 18. Correlation between the critical level or probability value of each test and the level of deviation from multivariate normality.

Correlated tests	r_s	SE	Fisher's transformation			95% CI for r_s	
			z_{rs}	LB_z	UB_z	LB	UB
Q' & D	−0.740	0.165	−0.951	−1.063	−0.417	−0.787	−0.395
Q & D	−0.746	0.165	−0.964	−1.069	−0.423	−0.789	−0.399
M & D	−0.645	0.160	−0.766	−0.959	−0.330	−0.744	−0.319
SJ & D	−0.455	0.153	−0.490	−0.755	−0.154	−0.638	−0.153
H & D	−0.664	0.161	−0.801	−0.980	−0.348	−0.753	−0.335
H' & D	−0.677	0.162	−0.823	−0.994	−0.360	−0.759	−0.345

Note: Q = the proposed Q-test from the Shapiro-Wilk W-statistics and the Q' test from the Shapiro-Francia W' statistics, M = the Mardia's K^2-test, SJ = the Smith-Jain runs Z-test, H = the Royston's H-test from the Shapiro-Wilk W statistics and the H' test from the Shapiro-Francia W' statistics, and D = the ordinal variable of deviation from multivariate normality, r_s = the value of Spearman's rank-order correlation or rho coefficient, $SE = \sqrt{(1+r_s^2)/(n-3)}$ = Bonett-Wright standard error for rho coefficient, z_{rs} = atanh(r_s) = the Fisher's hyperbolic arctangent transformation of r_s, $LB = z_{rs} - 1.96 \times EE$ = the lower bound and $UB = z_{rs} + 1.96 \times EE$ = the upper bound of the 95% confidence interval for the transformed correlation, 95% CI for r_s = 95% confidence interval for rho coefficient undoing the transformation: LB = tanh (LB_z) and UB = tanh (UB_z).

Table 19. Comparison of correlations between the six tests using the Steiger's Z-test.

T_1	T_2	T_1D	T_2D	T_1T_2	z	p	φ	r	i	a_c	Sig.	Dif.
Q'	Q	−0.740	−0.746	0.976	0.283	0.777	0.344	0.227	14	0.047	no	
Q'	M	−0.740	−0.645	0.646	−1.185	0.236	0.093	0.039	11	0.037	no	
Q'	SJ	−0.740	−0.455	0.503	−2.709	**0.007**	0.726	0.332	2	0.007	yes	*Q'D > SJD*
Q'	H	−0.740	−0.664	0.848	−1.377	0.168	0.076	0.159	8	0.027	no	
Q'	H'	−0.740	−0.677	0.846	−1.159	0.247	0.126	0.203	12	0.040	no	
Q	M	−0.746	−0.645	0.632	−1.248	0.212	0.094	0.094	10	0.034	no	
Q	SJ	−0.746	−0.455	0.531	−2.851	**0.004**	0.733	0.384	1	0.003	no	*QD > SJD*
Q	H	−0.746	−0.664	0.845	−1.482	0.138	0.078	0.064	6	0.020	no	
Q	H'	−0.746	−0.677	0.844	−1.266	0.205	0.133	0.112	9	0.030	no	
M	SJ	−0.645	−0.455	0.784	−2.447	**0.014**	0.768	0.404	3	0.010	no	*MD > SJD*
M	H	−0.645	−0.664	0.671	0.236	0.813	0.149	0.121	15	0.050	no	
M	H'	−0.645	−0.677	0.673	0.387	0.698	0.195	0.147	13	0.043	no	
SJ	H	−0.455	−0.645	0.484	1.642	0.101	0.805	0.379	5	0.017	no	
SJ	H'	−0.455	−0.677	0.467	1.937	0.053	0.838	0.396	4	0.014	no	
H	H'	−0.664	−0.677	0.997	1.383	0.167	0.352	0.221	7	0.024	no	

Note: rho_{T1D} = the Spearman's rank correlation coefficient between the first test (T_1) and the ordinal variable of deviation from multivariate normality (D), rho_{T2D} = the Spearman's rank correlation coefficient between the second test (T_2) and the ordinal variable of deviation from normality multivariate normality (D), rho_{T1T2} = the Spearman's rank correlation coefficient between the two tests (T_1 and T_2), z = the standardized value of the difference between correlations using the Steiger's formula [45], p = two-tailed probability under a standard normal distribution, a_c = level of significance with the Sidak's correction using the Holm's procedure, φ = statistical power (calculated with the GPower program for the Z-test of two dependent correlations with a common index: a = D, b = T_1, and c = T_2) [53], r = $|z|/\sqrt{n}$ = effect size measure, Sig = significance: no when $p \geq a_c$ and yes when $p < a_c$, and Diff. = higher correlation between T_1 or T_2 and D in each pairwise comparison.

4.5. On the Assumption of Independence in the 50 Tuples

Due to the small size of the sequences, a significance level of 10% was used. Among the 50 sequences of the 15 $z_{\ln(1-W')}$ values (from the Shapiro-Francia W' statistics), the ordinary Ljung-Box test detected serial dependence in multivariate samples 6, 12, 15, and 46. The robust Ljung-Box test from the sequences of z'_l values (transformed into ranks) detected serial dependence in multivariate samples 15, 22, 25, 38, and 46. Both tests agreed in samples 15 and 46. However, the ordinary Ljung-Box test applied to the reduced sequences (without zeros) did not confirm serial dependence in any of the seven cases. From the reduced sequences, there was serial dependence in the multivariate samples 39 (from four independent samples with chi-square distribution with 100 degrees of freedom) with a significant first-order lag autocorrelation ($ar_1 = -0.684 < LB_{90\%} = -0.672$) and 44 (from two independent samples with normal distribution and two independent samples with chi-square distribution with one degree of freedom) with a significant second-order lag autocorrelation ($ar_2 = -0.462 < LB_{90\%} = -0.457$). See Table 20.

Reviewing the 50 sequences of $z_{\ln(1-W)}$ values (from the Shapiro-Wilk W statistic), the assumption of independence was rejected by the ordinary Ljung-Box Q-test in multivariate samples 15 and 46. The robust Ljung-Box test from the sequences of z_l values (transformed into ranks) detected serial dependence in multivariate samples 6, 15, 22, 26, and 46. Once again, both tests coincided in the significance of samples 15 and 46. However, the ordinary Ljung-Box test did not confirm serial dependence in any of the five cases in the reduced sequences (without zeros). With the reduced sequences, the Ljung-Box Q-test was significant in the multivariate sample 29 of three correlated variables with normal distribution and one independent variable with exponential distribution. Its highest autocorrelation was that of first-order lag ($ar_1 = -0.666$). See Table 20.

Critical values were obtained by bootstrapping (Monte Carlo simulation) for multivariate sample 29 (proposed Q test) and multivariate samples 39 and 44 (proposed Q' test). Three, six, and thirteen standard half-normal distributions (truncated standard normal distribution between the 0.5 and 0.9999 quantiles) were defined, respectively. The outcome variable was the sum of squares of the 3, 6 or 13 variables with standard half-normal distributions. Three correlation matrices were defined with ones in the main diagonal, the value of the autocorrelation in the corresponding variables and zeros in the remaining cells. The correlation was -0.666 (first-order lag autocorrelation) between z_1 - z_2 and z_2 - z_3 in sample 29. The correlation was -0.684 (first-order lag autocorrelation) between z_1 - z_2, z_2 - z_3, z_3 - z_4, z_4 - z_5, and z_5 - z_6 in sample 39. The correlation was -0.462 (second-order lag autocorrelation) between z_1 - z_3, z_2 - z_4, z_3 - z_5, z_4 - z_6, z_5 - z_7, z_6 - z_8, z_7 - z_9, z_8 - z_{10}, z_9 - z_{11}, z_{10} - z_{12}, and z_{11} - z_{13} in sample 44. Correlations were estimated using Spearman's rank-order coefficient and calculations were performed with the XLSTAT software version 24 [20]. Percentiles were obtained from 1000 bootstrap samples for statistical decision making: if q' or $q >$ (simulated) 95th percentile, H_0 is rejected. As the autocorrelations were negative, the

Table 20. Testing the assumption of independence using the Ljung-Box Q test.

	Multivariate sample	Complete ($h = 3$)		Reduced		Complete ($h = 3$)		Reduced	
		$z_{\ln(1-W')}$	R(z'_l)	z'_l		$z_{\ln(1-W)}$	R(z_l)	z_l	
		p for LJ	p for LJr	h	p for LJ	p for LJ	p for LJr	h	p for LJ
1	$N_1E_1C_1L_1$	0.593	0.781			0.914	0.955		
2	$N_2E_2C_2L_2$	0.439	0.118			0.338	0.363		
3	$N_3E_3C_3L_3$	0.826	0.441			0.825	0.192		
4	$N_4E_4C_4L_4$	0.338	0.421			0.312	0.392		
5	$N_1E_2C_3L_4$	0.722	0.461			0.782	0.471		
6	**$N_1N_2N_3N_4$**	**0.087**	0.369	1	0.840	0.173	**0.051**	1	0.117
7	4N rel	0.734	0.279			0.631	0.942		
8	4E ind	0.878	0.807			0.894	0.940		
9	4C ind	0.545	0.773			0.616	0.688		
10	4L ind	0.632	0.464			0.687	0.369		
11	$N_1N_2E_1C_1$	0.156	0.146			0.189	0.150		
12	**$N_1N_2E_1L_1$**	**0.093**	0.120	2	0.683	0.119	0.174		
13	$N_1N_2C_1L_1$	0.527	0.512			0.373	0.622		
14	$N_1N_3E_2C_2$	0.216	0.191			0.162	0.155		
15	**$N_1N_3E_2L_2$**	**0.062**	**0.052**	2	0.710	**0.091**	**0.066**	2	0.215
16	$N_1N_3C_2L_2$	0.228	0.115			0.291	0.237		
17	$N_1N_4E_3C_3$	0.455	0.417			0.392	0.176		
18	$N_1N_4E_3L_3$	0.175	0.124			0.181	0.124		
19	$N_1N_4C_3L_3$	0.891	0.471			0.875	0.487		
20	$N_2N_3E_4C_4$	0.355	0.556			0.216	0.486		
21	$N_2N_3E_4L_4$	0.287	0.175			0.198	0.161		
22	$N_2N_3C_4L_4$	0.427	**0.047**	2	0.740	0.262	**0.028**	2	0.491
23	$N_2N_4E_1C_3$	0.939	0.672			0.868	0.598		
24	$N_2N_4E_1L_3$	0.680	0.644			0.583	0.612		
25	$N_2N_4C_1L_3$	0.833	0.479			0.869	0.539		
26	$N_3N_4E_2C_1$	0.456	**0.097**	3	0.721	0.361	**0.070**	2	0.289
27	$N_3N_4E_2L_1$	0.813	0.729			0.637	0.285		
28	$N_3N_4C_2L_1$	0.932	0.892			0.873	0.787		
29	3NE rel	0.654	0.227			0.884	0.602	**3**	**0.068**
30	3NC rel	0.478	0.172			0.252	0.308		
31	3NL rel	0.517	0.893			0.563	0.850		

Continued

32	4N rel	0.923	0.323			0.833	0.452		
33	4N rel	0.159	0.121			0.453	0.697		
34	4N rel	0.550	0.911			0.424	0.474		
35	4Logist	0.347	0.067			0.434	0.430		
36	4Lap ind	0.267	0.210			0.443	0.369		
37	4B(10,.1)	0.282	0.530			0.456	0.372		
38	4t(100)	0.746	**0.057**	2	0.344	0.326	0.221		
39	$4\chi^2(100)$	0.678	0.181	**6**	**0.034**	0.603	0.268		
40	4B (20,.5)	0.835	0.695			0.813	0.728		
41	4N ind	0.952	0.725			0.964	0.927		
42	$2N2N^{-1}$	0.317	0.186			0.286	0.186		
43	$2N2\chi^2_{2 \text{ or } 4}$	0.960	0.332			0.935	0.883		
44	$2N2\chi^2_1$	0.679	0.586	13	0.070	0.572	0.910		
45	2N2C	0.703	0.859			0.809	0.814		
46	**4N**	**0.005**	**0.014**	2	0.290	**0.018**	**0.078**	1	0.604
47	$3N1N^{-1}$	0.340	0.441			0.419	0.566		
48	$3N1\chi^2_2$	0.321	0.707			0.431	0.444		
49	$3N1\chi^2_1$	0.803	0.563			0.579	0.197		
50	3N1C	0.485	0.291			0.651	0.604		

Note: Complete sequence (sequence of 15 values): $z_{\ln(1-W')}$ = the standardized and log-transformed Shapiro-Francia W' statistics [4], z'_l = the truncated, standardized and log-transformed W' statistics, R(z'_l) = the range of the value z'_l with average ranks in case of a tie, $z_{\ln(1-W)}$ = the standardized and log-transformed Shapiro-Wilk W statistics [3], z_l = the truncated, standardized and log-transformed W statistics, R(z_l) = the range of the value z_l with average ranks in case of a tie. Reduced sequence: without the null z'_l or z_l values. p = asymptotic probability to the right tail in a chi-square distribution with three degrees of freedom in the complete sequence and with h degrees of freedom in the reduced sequence (without zeros), LJ = ordinary Ljung-Box Q-test, LJr = robust Ljung-Box Q-test, h = the order of the maximum lag determined using the Hyndman-Athanasopoulos rule [27]. Significant tests at a 10% significance level are highlighted in bold.

critical values or simulated percentiles were lower compared to the simulated percentiles with the independent variables or the percentiles of the chi-square distribution. See Table 21.

In samples 29 and 44, the null hypothesis of multivariate normality is rejected according to expectation (q = 6.922 < simulated P_{95} = 0.223 and q' = 45.9 < simulated P_{95} = 11.052, respectively). In sample 39, the null hypothesis is also rejected (q' = 5.322 < simulated P_{95} = 3.958), when the expectation is that it holds,

Table 21. Monte Carlo simulation quantiles.

p	Sample 29			Sample 39			Sample 44		
	$_p\chi_3^2$	Ind	Rel	$_p\chi_6^2$	Ind	Rel	$_p\chi_{13}^2$	Ind	Rel
0.05	0.352	0.347	0.008	1.635	1.602	0.464	5.892	5.743	2.929
0.10	0.584	0.610	0.014	2.204	2.201	0.625	7.042	7.189	3.666
0.20	1.005	1.045	0.023	3.070	3.130	0.858	8.634	8.573	4.372
0.25	1.213	1.242	0.028	3.455	3.470	0.959	9.299	9.219	4.702
0.30	1.424	1.492	0.033	3.828	3.876	1.085	9.926	10.100	5.151
0.40	1.869	1.897	0.044	4.570	4.572	1.319	11.129	11.172	5.698
0.50	2.366	2.344	0.057	5.348	5.324	1.504	12.340	12.594	6.423
0.60	2.946	3.008	0.071	6.211	6.193	1.736	13.636	13.626	6.949
0.70	3.665	3.667	0.090	7.231	7.326	2.053	15.119	15.014	7.657
0.75	4.108	4.040	0.102	7.841	8.009	2.230	15.984	15.927	8.123
0.80	4.642	4.673	0.116	8.558	8.655	2.457	16.985	16.928	8.633
0.90	6.251	6.249	0.162	10.645	10.586	3.169	19.812	19.412	9.900
0.95	7.815	7.537	0.223	12.592	12.743	3.958	22.362	21.671	11.052

Note: Sampling method: Latin hypercubes (number of sections = 500), number of intervals: 50, number of simulations: 1000, type of correlation: Spearman. $_p\chi_{df}^2$ = the p-order quantile of a chi-square distribution with df degrees of freedom, Ind = simulation with independent variables, and Rel = simulation with correlated variables.

so the test seems to perform better from the asymptotic approach by forcing the assumption of independence as is done in Mardia's K^2-test and multivariate runs Z-tests. Hence, it is advisable to use the reduced sequence only in case of significance in the full sequence. If the ordinary Ljung-Box test is used, the sequence must include the negative values. With the truncated sequence, the robust Ljung-Box test is used, which does not require normality and is more sensitive to anomalous tails [25] [26].

5. Discussion

The first objective of this article was to present a new multivariate normality test. It is based on the lemma or proven proposition that, if a set of correlated or independent variables follow a multivariate normal distribution, any linear combination of them follows a univariate normal distribution [14] [54]. If there are k variables, the number of unweighted linear combinations is $2^k - 1$. Additionally, the lemma that the sum of squares of $2^k - 1$ independent variables with standard normal distribution follows a chi-square distribution with $2^k - 1$ degrees of freedom is considered [15]. This proposal has two variants: one from the Shapiro-Wilk W statistics with the logarithmic transformation and standardization of

Royston [3] and the other from the Shapiro-Francia W' statistics with the logarithmic transformation and standardization of Royston [4]. Since the calculation of the critical level or probability value of the Royston´s statistics is one-sided (to the right tail), a problem arises with the negative values that indicate good fit. One option would be to take its absolute inverse, but this would result in a strong deviation in the sampling distribution of the test statistic from the chi-square distribution.

Another way of solving this problem was sought, and it was considered that the best option was to truncate these values to 0, so that the sampling distribution of the variables changes from a standard normal distribution to standard half-normal distribution, since truncating a standard normal variable between the quantiles 0.5 and 0. 9999... results in a standard half-normal variable: $x \in X \sim \text{SN}(\sigma = 1)$, $f(x) = 2 \times \varphi(x)$, $F(x) = 2 \times \Phi(x) - 1$, and $F^{-1}(x) = \Phi^{-1}[(p + 1)/2]$, where φ is the density function, Φ the distribution function and Φ^{-1} the quantile function of a standard normal distribution [17]. Two additional lemmas are added here. The first states that if a variable follows a half-normal distribution, the square of the quotient between the variable and its scale parameter σ follows a chi-squared distribution with one degree of freedom [17]. The second posits that the sum of $2^k - 1$ independent variables with chi-square distribution with one degree of freedom follows a chi-square distribution with $2^k - 1$ degrees of freedom [18]. As the scale parameter σ takes a unit value in the $2^k - 1$ variables, the sum of squares of each variable divided by σ is reduced to the sum of squares of the variables, thus returning to the starting point.

For the sum of squares of standard normal variables (of mean 0 and unit variance) or standard half-normal variables to follow a chi-square distribution with as many degrees of freedom as variables summed, independence between variables is required [18]. However, the test is based on all possible linear combinations among the k variables from which the Shapiro-Wilk W- or Shapiro-Francia W' statistics are calculated. Here one could object that they are dependent variables and that the sampling distribution is generalized chi-square, which is a very complex distribution to calculate [19]. Indeed, the generated variables are linearly dependent, but the z_l (from Shapiro-Wilk W statistics) or z_l' (from Shapiro-Francia W' statistics) values are not necessarily, so the independence assumption shifts to showing that the $2^k - 1$ z_l or z_l' values in their generative sequence are independent, *i.e.*, they do not exhibit serial correlation, either with a first-order lag or with a lag greater than one order. Here, these $2^k - 1$ values are conceived as identically distributed random variables with sample size 1. To test this assumption, the Ljung-Box Q test is used, and the maximum lag can be determined using the Hyndman-Athanasopoulos rule for non-stationary series [27]. The Ljung-Box test assumes bivariate normality in each autocorrelation, so it is required to use its robust version from the truncated sequence or to resort to the untruncated series. In case of non-compliance with the independence assumption, the critical values for the Q- or Q'-test can be obtained using

Monte Carlo simulation.

To improve the specificity of the test, an operational correction is introduced that consists of eliminating one degree of freedom for each canceled variable (negative value converted to zero). Consequently, the simulation is run with a simplified sequence that corresponds to the non-zero values (random variables with sample size 1). Hence, the serial independence test has to be repeated with the simplified sequence, and from there obtain the significant autocorrelation values for the simulation. To obtain these last values, the correlogram is a very useful tool [23]. If there is no serial dependency in this second test, an ambiguous situation is generated that is resolved in favor of independence. One can also go directly to the simplified sequence. However, it seems that it is better to force the assumption of independence, so this second path is not recommended. It is only considered that there is a serial dependency if it appears in both the complete and reduced sequence with some significant autocorrelation.

The second objective of the study sought to compare the central tendency of the type II error or β-probability and the statistical power or complement of the β-probability. From these central tendency analyses, the new test yields good results without a clear advantage of one of its two versions, being equivalent to the Royston's test and superior to the Mardia's K^2-test and runs Z-test. The third objective involved to compare the frequency of successes among the six multivariate normality tests. In this analysis, the proposed test together with the Royston's test are the best, with no difference between their two versions. Although there is a situation of ambiguity in the statistical decision of equivalence between the two versions of the proposed test. The difference is not significant, but the effect size is large, the type II error is null and the unitary statistical power, which supports the alternative hypothesis of difference, in which the Q-test would have a higher proportion of successes than the Q'-test, 0.94 versus 0.88. The fourth objective focused on calculating and comparing the sensitivity, specificity, and efficiency of the six multivariate normality tests. All six tests have unit sensitivity, so they are equivalent in the property of detecting normality. Comparisons in specificity and efficiency among the six tests do not reveal significant differences, even without controlling for family rate error, so it would seem that average values would be valid for all of them and these would be high, namely above 0.85. However, when testing higher-than-average specificity and efficiency values (H_0: $E \geq 0.90$ and H_0: $Ef \geq 0.90$), the runs Z-test and Mardia's K^2-test show significantly lower specificity values than those hypothesized. The fifth objective proposed to classify the samples into ordered categories of deviation from normality, calculate the correlation between this ordinal variable and the critical level or probability value of each of the six tests, and compare these correlations. The highest correlations appear in the proposed test and the lowest appears with the runs Z-test. The latter presents a significant difference only in comparison with the two versions of the proposed test. In this analysis, the proposed test stands out with no difference between its two versions.

6. Conclusions

The new proposal presents a performance very similar to Royston's test and clearly superior to the Mardia's K^2-test and runs a Z-test with samples of 20 tetra-dimensional tuples. It should be noted that the Q version from the Shapiro-Wilk W statistics reveals a very slight advantage over the Q' version from the Shapiro-Francia W' statistics. In the face of samples of 20 participants and 4 variables, the highest specificity and efficiency values, success ratio, and correlation between the critical level and the ordinal variable of deviation from normality are achieved with the proposed Q-test.

As limitations of the study, it should be noted that the number of simulations is very small, so it is merely a pilot study. With a larger number of simulations, some of the differences between the two versions of the proposed test may be significant, and the version based on the Shapiro-Wilk W statistics may be more specific and efficient. This pilot study only handles one sample size ($n = 20$) and one number of variables ($k = 4$) when the variation of n and k would allow the definition of power curves to compare the tests. There are also other multivariate normality tests not covered in the present work, such as those of Cox and Small [55], Henze and Zirkler [56], and Doornik and Hansen [57], available in the R program [58], the Monte Carlo version of multivariate runs test available in Excel [9] [59] or the tests of Arnastauskaite, Ruzgas and Braženas [60] and Kesemen, Tiryaki, Tezel and Özkul [61], more recently. There is also another generalization of the Shapiro-Wilk test developed by Villaseñor-Alva and González-Estrada [62], different from that of Royston [10] and the present work. Any of these tests not included would be excellent comparison options for future research, although there is currently no evidence or consensus on which is the best test [58] [60].

Further study of this new statistical test is suggested. A test based on the principle that, given k variables drawn from a multivariate normal population, any linear combination of these variables should follow a univariate normal distribution; additionally, on the principle that the sum of squares of independent standard half-normal variables follows a chi-square distribution with as many degrees of freedom as variables added, with the additional correction of eliminating the number of variables nulled in the degrees of freedom. The independence assumption is tested on the generative sequence of standardized and truncated values. Initially, it is checked with the complete series and, in case of dependence, it is repeated with the reduced series (without zeros). If both series show dependence, the assumption is not fulfilled. In case of discrepancy between the independence of the two series, the decision is in favor of independence, since the simulated quantiles with moderate or high negative correlations go down a lot or with moderate or high positive correlations go up a lot compared to the situation with independent samples, resulting in a less accurate test. If further studies support this test, its computational implementation in programs such as R or Excel is very simple. Precisely, this article details an example, executed from

the Excel program.

Acknowledgements

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1]   Shapiro, S.S. and Wilk, M.B. (1965) An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, **52**, 591-611. https://doi.org/10.1093/biomet/52.3-4.591

[2]   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. https://doi.org/10.1080/01621459.1972.10481232

[3]   Royston, J.P. (1992) Approximating the Shapiro-Wilk W-Test for Non-Normality. *Statistics and Computing*, **2**, 117-119. https://doi.org/10.1007/BF01891203

[4]   Royston, J.P. (1993) A Tool Kit for Testing for Normality in Incomplete and Censored Samples. *Journal of the Royal Statistical Society. Series D* (*the Statistician*), **42**, 37-43. https://doi.org/10.2307/2348109

[5]   Mardia, K.V. (1970) Measures of Multivariate Skewness and Kurtosis with Applications. *Biometrika*, **57**, 519-530. https://doi.org/10.1093/biomet/57.3.519

[6]   Mardia, K.V. (1974) Applications of Some Measures of Multivariate Skewness and Kurtosis in Testing Normality and Robustness Studies. *Sankhya: The Indian Journal of Statistics*, *Series B* (1960-2002), **36**, 115-128. https://www.jstor.org/stable/25051892

[7]   Mardia, K.V. (1980) Tests of Univariate and Multivariate Normality. In Krishnaiah, P.R., Ed., *Handbook of Statistics* 1: *Analysis of Variance*, North-Holland, Amsterdam, 279-320. https://doi.org/10.1016/S0169-7161(80)01011-5

[8]   Friedman, J.H. and Rafsky, L.C. (1979) Multivariate Generalizations of the Wald-Wolfowitz and Smirnov Two-Sample Tests. *The Annals of Statistics*, **7**, 697-717. https://doi.org/10.1214/aos/1176344722

[9]   Smith, S.P. and Jain, A.K. (1988) A Test to Determine the Multivariate Normality of a Data Set. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **10**, 757-761. https://doi.org/10.1109/34.6789

[10] Royston, J.P. (1983) Some Techniques for Assessing Multivariate Normality Based on the Shapiro-Wilk W. *Journal of the Royal Statistical Society. Series C* (*Applied Statistics*), **32**, 121-133. https://doi.org/10.2307/2347291

[11] Wald, A. (1939) Contributions to the Theory of Statistical Estimation and Testing Hypotheses. *Annals of Mathematical Statistics*, **10**, 299-326. https://doi.org/10.1214/aoms/1177732144

[12] Steiger, J.H., Shapiro, A. and Browne, M.W. (1985) On the Multivariate Asymptotic Distribution of Sequential Chi-Square Statistics. *Psychometrika*, **50**, 253-263. https://doi.org/10.1007/BF02294104

[13] Hoaglin, D.C. (2016) Misunderstandings about Q and Cochran's Q Test. In Meta-Analysis. *Statistics in Medicine*, **35**, 485-495. https://doi.org/10.1002/sim.6632

[14] Ghurye, S.G. and Olkin, I. (1962) A Characterization of the Multivariate Normal Distribution. *The Annals of Mathematical Statistics*, **33**, 533-541. https://doi.org/10.1214/aoms/1177704579

[15] Cochran, W.G. (1934) The Distribution of Quadratic Forms in a Normal System, with Applications to the Analysis of Covariance. *Mathematical Proceedings of the Cambridge Philosophical Society*, **30**, 178-191. https://doi.org/10.1017/S0305004100016595

[16] Blom, G. (1958) Statistical Estimates and Transformed Beta-Variables. John Wiley and Sons, New York.

[17] Johnson, N., Kotz, S. and Balakrishnan, N. (1994) Continuous Univariate Distributions. 2nd Edition, John Wiley and Sons, New York.

[18] Coelho, C.A. (2020) On the Distribution of Linear Combinations of Chi-Square Random Variables. In Bekker, A., Chen, D.G. and Ferreira, J.T., Eds., *Computational and Methodological Statistics and Biostatistics. Emerging Topics in Statistics and Biostatistics*, Springer, Cham, 211-250. https://doi.org/10.1007/978-3-030-42196-0_9

[19] Rahman, G., Mubeen, S. and Rehman, A. (2015) Generalization of Chi-Square Distribution. *Journal of Statistics Applications and Probability*, **4**, 119-126. https://digitalcommons.aaru.edu.jo/jsap/vol4/iss1/12

[20] Addinsoft (2021) Monte Carlo Simulations. In XL-STAT: Tutorials & Guides. https://help.xlstat.com/tutorial-guides/monte-carlo-simulations

[21] Wald, A. and Wolfowitz, J. (1943) An Exact Test for Randomness in the Non-Parametric Case Based on Serial Correlation. *Annals of Mathematical Statistics*, **14**, 378-388. https://doi.org/10.1214/aoms/1177731358

[22] Ljung, G.M. and Box, G.E.P. (1978) On a Measure of a Lack of Fit in Time Series Models. *Biometrika*, **65**, 297-303. https://doi.org/10.1093/biomet/65.2.297

[23] Box, G.E.P., Jenkins, G.M., Reinsel, G.C. and Ljung, G.M. (2015) Time Series Analysis: Forecasting and Control. 5th Edition, John Wiley and Son, New York.

[24] Uyanto, S.S. (2020) Power Comparisons of Five Most Commonly Used Autocorrelation Tests. *Pakistan Journal of Statistics and Operation Research*, **16**, 119-130. https://doi.org/10.18187/pjsor.v16i1.2691

[25] Chan, W.S. (1994) On Portmanteau Goodness-of-Fit Tests in Robust Time Series Modeling. *Computational Statistics*, **9**, 301-310.

[26] Burns, P.J. (2002) Robustness of the Ljung-Box Test and Its Rank Equivalent. *SSRN*. https://doi.org/10.2139/ssrn.443560

[27] Hyndman, R.J. and Athanasopoulos, G. (2021) Forecasting: Principle and Practice. 3rd Edition, OTexts, Melbourne.

[28] Schwert, G.W. (1989) Why Does Stock Market Volatility Change Over Time? *Journal of Finance*, **44**, 1115-1153. https://doi.org/10.1111/j.1540-6261.1989.tb02647.x

[29] Albuquerque, P. (2020) Optimal Time Interval Selection in Long-Run Correlation Estimation. *Journal of Quantitative Economics, the Indian Econometric Society*, **18**, 53-79. https://doi.org/10.1007/s40953-019-00175-x

[30] Cohen, J. (1988) Statistical Power Analysis for the Behavioral Sciences. 2nd Edition, Lawrence Erlbaum Associate, Hillsdale.

[31] Friedman, M. (1937) The Use of Ranks to Avoid the Assumption of Normality Implicit in Analysis of Variance. *Journal of the American Statistical Association*, **32**, 675-701. https://doi.org/10.1080/01621459.1937.10503522

[32] Kendall, M.G. and Gibbons, J.D. (1990) Rank Correlation Methods. 5th Edition, A.

Charles Griffin, London.

[33] Wilcoxon, F. (1945) Comparison by Ranking Methods. *Biometrics Bulletin*, **1**, 80-83. https://doi.org/10.2307/3001968

[34] Rosenthal, R. (1991) Metanalytic Procedures for Social Research. Rev. Edition, Sage Publications, Inc., New York.

[35] Sidak, Z. (1967) Rectangular Confidence Regions for the Means of Multivariate Normal Distributions. *Journal of the American Statistical Association*, **62**, 626-633. https://doi.org/10.2307/2283989

[36] Holm, S. (1979) A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Statistical Journal*, **6**, 65-70.

[37] Cochran, W.G. (1950) The Comparison of Percentages in Matched Samples. *Biometrika*, **37**, 256-266. https://doi.org/10.1093/biomet/37.3-4.256

[38] Serlin, R.C., Carr, J. and Marascuilo, L.A. (1982) A Measure of Association for Selected Nonparametric Procedures. *Psychological Bulletin*, **92**, 786-790. https://doi.org/10.1037/0033-2909.92.3.786

[39] McNemar, Q. (1947) Note on the Sampling Error of the Difference between Correlated Proportions and Percentages. *Psychometrika*, **12**, 153-157. https://doi.org/10.1007/BF02295996

[40] Wilson, E.B. (1927) Probable Inference, the Law of Succession, and Statistical Inference. *Journal of the American Statistical Association*, **22**, 209-212. https://doi.org/10.1080/01621459.1927.10502953

[41] Newcombe, R.G. (1998) Two-Sided Confidence Intervals for the Single Proportion: Comparison of Seven Methods. *Statistics in Medicine*, **17**, 857-872. https://doi.org/10.1002/(SICI)1097-0258(19980430)17:8<857::AID-SIM777>3.0.CO;2-E

[42] Spearman, C. (1904) The Proof and Measurement of Association between Two Things. *The American Journal of Psychology*, **15**, 72-101. https://doi.org/10.2307/1412159

[43] Fisher, R.A. (1915) Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population. *Biometrika*, **10**, 507-521. https://doi.org/10.1093/biomet/10.4.507

[44] Bonett, D.G. and Wright, T.A. (2000) Sample Size Requirements for Estimating Pearson, Kendall and Spearman Correlations. *Psychometrika*, **65**, 23-28. https://doi.org/10.1007/BF02294183

[45] Steiger, J.H. (1980) Tests for Comparing Elements of a Correlation Matrix. *Psychological Bulletin*, **87**, 245-251. https://doi.org/10.1037/0033-2909.87.2.245

[46] Myers, L. and Sirois, M.J. (2006) Spearman Correlation Coefficients, Differences Between. In: Kotz, S., Balakrishnan, N., Read, C.B. and Vidakovic, B., Eds., *Encyclopedia of Statistical Sciences*, 2nd Edition, John Willey and Sons, Hoboken, 7901-1903. https://doi.org/10.1002/0471667196.ess5050.pub2

[47] D'Agostino, R.B., Belanger, A. and D'Agostino Jr., R.B. (1990) A Suggestion for Using Powerful and Informative Tests of Normality. *The American Statistician*, **44**, 316-321. https://doi.org/10.1080/00031305.1990.10475751

[48] Verma, J.P. and Abdel-Salam, A.S.G. (2019) *Testing Statistical Assumptions in Research*. John Wiley and Sons, Newark. https://doi.org/10.1002/9781119528388

[49] Chakraborti, S. and Graham, M.A. (2019) Nonparametric (Distribution-Free) Control Charts: An Updated Overview and Some Results. *Quality Engineering*, **31**, 523-544. https://doi.org/10.1080/08982112.2018.1549330

[50] D'Agostino, R.B. (1970) Transformation to Normality of the Null Distribution of g_1.

Biometrika, **57**, 679-681. https://doi.org/10.1093/biomet/57.3.679

[51] Anscombe, F.J. and Glynn, W.J. (1983) Distribution of the Kurtosis Statistic b$_2$ for Normal Samples. *Biometrika*, **70**, 227-234. https://doi.org/10.1093/biomet/70.1.227

[52] Jovanovic, B.D. and Levy, P.S. (1997) A Look at the Rule of Three. *The American Statistician*, **51**, 137-139. https://doi.org/10.1080/00031305.1997.10473947

[53] Universität Düsseldorf, Psychologie (2021) G* Power 3.1 Manual. https://www.psychologie.hhu.de/fileadmin/redaktion/Fakultaeten/Mathematisch-Naturwissenschaftliche_Fakultaet/Psychologie/AAP/gpower/GPowerManual.pdf

[54] Rao, C.R. (1973) Linear Statistical Inference and Its Applications. 2nd Edition, John Wiley and Sons, New York. https://doi.org/10.1002/9780470316436

[55] Cox, D.R. and Small, N.J.H. (1978) Testing Multivariate Normality. *Biometrika*, **65**, 263-272. https://doi.org/10.1093/biomet/65.2.263

[56] Henze, N. and Zirkler, B. (1990) A Class of Invariant Consistent Tests for Multivariate Normality. *Communications in Statistics—Theory and Methods*, **19**, 3595-3617. https://doi.org/10.1080/03610929008830400

[57] Doornik, J.A. and Hansen, H. (2008) An Omnibus Test for Univariate and Multivariate Normality. *Oxford Bulletin of Economics and Statistics*, **70**, 927-939. https://doi.org/10.1111/j.1468-0084.2008.00537.x

[58] Korkmaz, S. (2022) Package "MVN". Multivariate Normality Tests. https://cran.r-project.org/web/packages/MVN/MVN.pdf

[59] Zaiontz, C. (2022) Multivariate Normality Testing (FRSJ). In Real Statistics Using Excel. https://www.real-statistics.com/multivariate-statistics/multivariate-normal-distribution/multivariate-normality-testing-frsj

[60] Arnastauskaite, J., Ruzgas, T. and Braženas, M.A (2021) New Goodness of Fit Test for Multivariate Normality and Comparative Simulation Study. *Mathematics*, **9**, Article No. 3003. https://doi.org/10.3390/math9233003

[61] Kesemen, O., Tiryaki, B.K., Tezel, Ö. and Özkul, E. (2021) A New Goodness of Fit Test for Multivariate Normality. *Hacettepe Journal of Mathematics and Statistics*, **50**, 872-894. https://doi.org/10.15672/hujms.644516

[62] Villaseñor-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**, 1870-1883. https://doi.org/10.1080/03610920802474465