ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 225, 2009

*Czesław Domański**

ATTEMPT TO ASSESS MULTIVARIATE NORMALITY TESTS

Abstract

The assumption of multivariate normality is a basis of the classical multivariate statistical methodology. Consequences of departures from these assumptions have not been investigated well so far. There are many methods of constructing multivariate normality tests. Some of these tests are broader versions of univariate normality tests.

Most of the multivariate normality tests which can be found in literature, can be divided into four categories:

1. Graph based procedures.
2. Generalized goodnes-of-fit tests.
3. Tests based of skewness and kurtosis measures.
4. Procedures based on empirical characteristic function.

The present paper is an attempt to assess selected tests from the point of view of their properties as well as possibilities of their applications.

Key words: multivariate normality tests, critical values, Shapiro-Wilk test.

1. Introduction

An overview of the subject literature shows that there exist at least 50 procedures for testing multivariate normality. Despite the multitude of methods, Rencher (1995) noted that since multivariate normality is not as straightforward as univariate normality, the "state of the art" is not so refined. Although several reviews of different methods were prepared (including Andrews, Gnanadesikan, Warner, 1977; Gnanadesikan, 1977; Koziol, 1986; Looney, 1995; Henze, 2002; Mudholkar and Srivastava, 2002), none of them, however, is fully comprehensive. What is more, permanent proliferation of papers containing new methods for testing multivariate normality makes it impossible to cover all available tests. Taking

* Professor, Chair of Statistical Methods, University of Łódź.

It is widely known that (see e.g. Bilodeau and Brenner, 1999) a data vector is distributed as a p-dimensional multivariate normal distribution if and only if all linear combinations of this vector are univariate normal, i.e.

$$x \sim N_p(\mu,\Sigma) \Leftrightarrow \mathbf{t}'\mathbf{x} \sim \acute{N}(\mathbf{t}'\mu,\mathbf{t}'\Sigma), \forall \mathbf{t} \in \mathbb{R}^n \quad (1)$$

It may seem natural that this fact is exploited by testing for multivariate normality only by testing a linear combination for univariate normality, where we can apply a well-regarded test of univariate normality. However, even if we are able to find a univariate normal lineal combination, we cannot state that the data vector is univariate normal unless we can prove that all such linear combinations are univariate normal.

2. Multivariate normality hypotheses

Let $\mathbf{x}_1, \dots \mathbf{x}_n$ be independent p-dimensional random vectors of an identical distribution defined by a distribution function $F_p(\mathbf{x})$, where $\mathbf{X} \in \Re^p$, and $\Re^p$ are p-dimensional real space. Let us denote a sample of $n(n > p)$ p-dimensional observations by means of the following matrix:

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1n} \\ X_{21} & X_{22} & \dots & X_{2n} \\ \dots & \dots & \dots & \dots \\ X_{p1} & X_{p2} & \dots & X_{pn} \end{bmatrix} = [\mathbf{x}_1, \dots, \mathbf{x}_n] \quad (2)$$

while we apply the same denotations for random vectors as well as their realizations. Let us denote p-dimensional random vector by $\mathbf{x}$ and let $N_p(\mathbf{x};\mu,\Sigma)$ be a distribution function of p-dimensional normal distribution, where $x \in \Re^p$ is a point of $\Re^p$ and μ and Σ are p-dimensional vectors of an expected value and $(p \times p)$-dimensional covariance matrix (dispersion matrix), respectively. The fact that the random vector $\mathbf{x}$ has a distribution determined by the distribution function $N_p(\mathbf{x};\mu,\Sigma)$ can be symbolically expressed as follows: $x \sim \Phi_p,(\mu,\Sigma)$.

Let us assume that *HSMN* (simple hypothesis multivariate normality) denotes a hypothesis of the form:

$$HSMN: F_p(x) = N_p(x;\mu_0,\Sigma_0) \quad (3)$$

into account the amount of work done in developing these tests, relatively little has been done as far as evaluation of the quality and power of testing procedures is concerned.

Mudholkar and Srivastava (2002) in their short paper focused on discussing a small number of the possible tests and contradictory results that can be obtained by applying different tests of multivariate normality in well-known data sets. They concluded that the assumption of multivariate normality is "illusory" and that the attention should be drawn to understanding the effect of non-normality and developing robust procedures of data analysis.

This paper aims mainly at identifying procedures available in the literature and describing in more detail some of the procedures which seem to be most promising from the point of view of invariance and consistency. Moreover, the author considers the potential which can be used by researchers, even those without a strong statistical background. The final condition leads us to the tests whose critical values have a well-known asymptotic null distribution which can be easily determined, to the detriment of those procedures which require critical values that can be read from specialized tables or determined by means of simulations carried out by the user.

Many of multivariate normality test procedures are extensions of tests of univariate normality. Most of the available tests of multivariate probability can be devided into four categories:

- procedures based on graphical plots and correlation coefficients,
- goodness-of-fit tests,
- tests based on skewness and kurtosis measures,
- consistent procedures based on the empirical characteristic function.

Some of the tests do not fit any of the categories, mainly because they are conceived as tests that can be used only in special circumstances, such as against a particular type of alternative distribution or with a particular form of data.

A lot of tests are used to test goodness-of-fit of the univariate normal distribution, including the well-regarded Shapiro-Wilk and Kolmogorov-Smirnov tests. If we assume that this approach could establish, by means of the rejection of univariate normality of at least one component, that a random data vector is not multivariate normal, then we can note that it does not do anything towards showing fit to the multivariate normal. It is possible for a multivariate distribution to have each univariate marginal distribution that is normal without joint normality. Bilodeau and Brenner (1999) give two examples to support the above statement; one where marginal distributions are normal but the joint distribution is a mixture of multivariate normal distributions and the other where the two marginal distributions are normal but the joint distribution is a Frank density rather than a multivariate normal distribution.

It is widely known that (see e.g. Bilodeau and Brenner, 1999) a data vector is distributed as a p-dimensional multivariate normal distribution if and only if all linear combinations of this vector are univariate normal, i.e.

$$x \sim N_p(\mu, \Sigma) \Leftrightarrow \mathbf{t}'\mathbf{x} \sim \acute{N}(\mathbf{t}'\mu, \mathbf{t}'\Sigma), \forall \mathbf{t} \in \mathbb{R}^n \quad (1)$$

It may seem natural that this fact is exploited by testing for multivariate normality only by testing a linear combination for univariate normality, where we can apply a well-regarded test of univariate normality. However, even if we are able to find a univariate normal lineal combination, we cannot state that the data vector is univariate normal unless we can prove that all such linear combinations are univariate normal.

2. Multivariate normality hypotheses

Let $\mathbf{x}_1, \dots \mathbf{x}_n$ be independent p-dimensional random vectors of an identical distribution defined by a distribution function $F_p(\mathbf{x})$, where $\mathbf{X} \in \Re^p$, and $\Re^p$ are p-dimensional real space. Let us denote a sample of $n (n > p)$ p-dimensional observations by means of the following matrix:

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1n} \\ X_{21} & X_{22} & \dots & X_{2n} \\ \dots & \dots & \dots & \dots \\ X_{p1} & X_{p2} & \dots & X_{pn} \end{bmatrix} = [\mathbf{x}_1, \dots, \mathbf{x}_n] \quad (2)$$

while we apply the same denotations for random vectors as well as their realizations. Let us denote p-dimensional random vector by $\mathbf{x}$ and let $N_p(\mathbf{x}; \mu, \Sigma)$ be a distribution function of p-dimensional normal distribution, where $x \in \Re^p$ is a point of $\Re^p$ and μ and Σ are p-dimensional vectors of an expected value and $(p \times p)$-dimensional covariance matrix (dispersion matrix), respectively. The fact that the random vector $\mathbf{x}$ has a distribution determined by the distribution function $N_p(\mathbf{x}; \mu, \Sigma)$ can be symbolically expressed as follows: $x \sim \Phi_p, (\mu, \Sigma)$.

Let us assume that *HSMN* (simple hypothesis multivariate normality) denotes a hypothesis of the form:

$$HSMN : F_p(x) = N_p(x; \mu_0, \Sigma_0) \quad (3)$$

i.e. $F_p(\mathbf{x})$ is a distribution function of the distribution $N_p(\mu_0,\Sigma_0)$, where μ_0 and Σ_0 are the parameters concerned. In particular, *HSMN** denotes a hypothesis of the form:

$$HSMN^*: F_p(\mathbf{x}) = N_p(\mathbf{x};\mathbf{0},\mathbf{I}) \tag{4}$$

that is $F_p(\mathbf{x})$ is a distribution function of the distribution $N_p(\mathbf{0};\mathbf{I})$.

Let, moreover, *HCMN* (composite hypothesis multivariate normality) denote a composite hypothesis:

$$HSMN^*: F_p(x) = N_p(\mathbf{x};\mu,\Sigma), \tag{5}$$

that is $F_p(\mathbf{x})$ is a distribution function of the distribution $N_p(\mu;\Sigma)$ of unknown parameters μ and Σ.

Unbiased estimators of parameters μ and Σ, obtained by means of a generalized least random squares method, are a sampling vector of expected values $\overline{\mathbf{x}}$ and a sampling matrix $\mathbf{S}$, respectively. They are as follows:

$$n\overline{\mathbf{x}} = \mathbf{x1} = \sum_{j=1}^{n} \mathbf{x}_j \tag{6}$$

where $\mathbf{1}$ is n-dimensional vector consisting of ones, and:

$$(n-1)\mathbf{S} = \mathbf{x}\left(\mathbf{I} - \frac{1}{n}\mathbf{11}^T\right)\mathbf{x}^T = \sum_{j=1}^{n}(\mathbf{x}_j - \overline{\mathbf{x}})(\mathbf{x}_j - \overline{\mathbf{x}}_j)^T = \sum_{j=1}^{n}\mathbf{x}_j\mathbf{x}_j^T - n\overline{\mathbf{x}\mathbf{x}}^T.$$

Both estimators $\overline{\mathbf{x}}$ and S are mutually stochastically independent and form a configuration of statistics sufficient for μ and Σ. What is more, if $\mathbf{x} \sim N(\mu;\Sigma)$, then $\overline{\mathbf{x}} \sim N_p\left(\mu;\frac{1}{n}\Sigma\right)$ and $\mathbf{S}$ has p-dimensional non-central Wishart distribution $W_p(n;\Sigma,\mu)$. The sampling covariance matrix is determined non-negatively, although in most practical cases it is determined positively.

If the matrix Σ is determined positively, then there exists such a matrix $\Sigma^{\frac{1}{2}}$ being a symmetrical square root of the matrix Σ, that $\Sigma^{-\frac{1}{2}}\Sigma\Sigma^{-\frac{1}{2}} = \mathbf{I}$. Then for the random vector $\mathbf{y} = \Sigma^{-\frac{1}{2}}(\mathbf{x} - \mu)$ we obtain $E(\mathbf{y}) = \mathbf{0}$ and $D(\mathbf{y}) = \mathbf{I}$. ($E(\cdot)$ and $D(\cdot)$ denote an expected value, a variance and a covariance matrix, respectively). Therefore, if $\mathbf{x} \sim N_p(\mu;\Sigma)$, then $y \sim N_p(\mathbf{0};\mathbf{I})$.

If we assume that a rank of the matrix Σ equals $(k < p)$, then we can find a matrix Σ^+ being a generalized Moor-Penros inverse of the matrix Σ and $(\Sigma^+)^{\frac{1}{2}} = (\Sigma^{\frac{1}{2}})^+$. It is possible since $\Sigma = \mathbf{P}\mathbf{D}_\lambda\mathbf{P}^T$ and $\Sigma^{\frac{1}{2}} = \mathbf{P}\mathbf{D}_\lambda^{\frac{1}{2}}\mathbf{P}^T$, where $\mathbf{P}$ is an orthogonal matrix whereas $\mathbf{D}_\lambda$ is a diagonal matrix of eigenvalues of the matrix Σ on a diagonal. At the same time, parameters of a random vector distribution

$$\mathbf{z} = (\Sigma^{\frac{1}{2}})^+(\mathbf{x} - \mu) \text{ are } E(\mathbf{z}) = \mathbf{0} \text{ and}$$

$$D(z) = (\Sigma^{\frac{1}{2}})^+ \Sigma[(\Sigma^{\frac{1}{2}})^+]^T = \mathbf{P}\mathbf{D}_\lambda^{-\frac{1}{2}}\mathbf{P}^T\mathbf{P}\mathbf{D}_\lambda\mathbf{P}^T\mathbf{P}\mathbf{D}_\lambda^{-\frac{1}{2}}\mathbf{P}^T =$$
$$= \mathbf{P}\mathbf{D}_\lambda^{-\frac{1}{2}}\mathbf{D}_\lambda\mathbf{D}_\lambda^{-\frac{1}{2}}\mathbf{P}^T = \mathbf{P}\mathbf{P}^T = \mathbf{I}$$

Therefore, if $x \sim N_p(\mu;\Sigma)$ and $r(\Sigma) = k < p$, to $\mathbf{z} \sim N_k(\mathbf{0};\mathbf{I})$.

The further part of this paper presents testing procedures for verification of hypotheses that the sample $\mathbf{X}$ stems from a multivariate normal population. Still, we assume that $rz(\Sigma) = p$, if $rz(\Sigma) = k$, then in the particular tests it is necessary to replace p with k (see Mardia, 1980).

The union-intersection principle of Roy (1953) was used by Malkovich and Afifi (1973) to generalize some of the univariate normality tests for a multivariate case. They based their assumptions on the theorem that $\mathbf{x} \sim N_p(\mu;\Sigma)$ if and only if $\mathbf{c}^T\mathbf{x} \sim N_1(\mathbf{c}^T\mu, \mathbf{c}^T\Sigma\mathbf{c})$ for all vectors $\mathbf{c} \in \Re^p$ and $\mathbf{c} \neq \mathbf{0}$. The proof for this theorem can be found, among others, in a monograph of Rao (1982).

Let us denote by $\mu_k(\mathbf{c}) = E\left\{\left[\mathbf{c}^T\mathbf{x} - \mathbf{c}^T E(\mathbf{x})\right]^k\right\}$ a central moment of the rank k, $(k = 2, 3, ...)$ dependent on a vector $\mathbf{c} \neq 0$. Malkovich and Afifi (1973) proposed measures of distributions of the following form:

a) multivariate skewness coefficient:

$$\beta_{1,p}(\mathbf{c}) = \mu_3^2(\mathbf{c}) / \mu_2^2(\mathbf{c}) \tag{7}$$

b) multivariate kurtosis coefficient:

$$\beta_{2,p}(\mathbf{c}) = \mu_4(\mathbf{c}) / \mu_2^2(\mathbf{c}) \tag{8}$$

Their sampling counterparts are, respectively:

$$b_{1,p}(\mathbf{c}) = nm_3^2 / m_2^3 \text{ i } b_{2,p}(\mathbf{c}) = nm_4 / m_2^2 \tag{9}$$

where:

$$m_k = \frac{1}{n}\sum_{j=k}^{n}(L_j - \overline{L})^k \quad (k = 2, 3, ...)$$

is a sample central moment $L_1, ... L_n$ with an average $\overline{L}$, while

$$L_j = \mathbf{c}^T\mathbf{x}_j \quad (j = 1, ..., n).$$

There is no reason to reject the hypothesis $H_0 : \beta_{1,p}(\mathbf{c}) = 0$, if $b_{1,p}^* = \max_{C \neq 0} b_{1,p}(\mathbf{c}) \leq K_{b_1}$, where K_{b_1} is a certain constant. Similarly, there is no reason to reject the hypothesis $H_0' : \beta_{2,p}(\mathbf{c}) = 0$, if

$(b_{2,p}^*)^2 = \max_{C \neq 0}\{b_{2,p}(\mathbf{c}) - K(n)\}^2 \leq K_{b_2}$, where $K(n) \to 3$, if $n \to \infty$ and K_{b_2} is a certain constant. The constants K_{b_1}, K_{b_2}, taking into consideration the particular conditions imposed on $\mathbf{c}$ (e.g. $\mathbf{c}^T\mathbf{c} = 1$) can be found by means of the Monte Carlo method.

Ordering L_j in a non-decreasing sequence $L_{(1)} \leq ... \leq L_{(n)}$ we express a Shapiro-Wilk (1965) statistic towards a vector $\mathbf{c}$ of the form:

$$W(c) = \left[\sum_{j=1}^{n} a_{j,n} L(j)\right]^2 / (n \cdot m_2) \quad (10)$$

where $\{a_{j,n}\}$ are constant normalized coefficients (applied in the paper of, among others, Domański (1990)), satisfying the following conditions:

$$\sum_{j=1}^{n} a_{j,n} = 0 \text{ and } \sum_{j=1}^{n} a_{j,n}^2 = 1 \quad (n = 3, 4, ...) \quad (11)$$

Starting with a property of the statistic $\mathbf{W(C)}$ for univariate normality, Malkovich and Afifi (1973) proposed choosing $\mathbf{c} = \mathbf{A}^{-1}(\mathbf{x}_1 - \overline{\mathbf{x}}) / a_{1,n}$, where $\mathbf{A} = (n-1)\mathbf{S}$. Then, in order to obtain the value $W(\mathbf{c})$, it s necessary to take the vector $\mathbf{x}_1$, that is one from among $\mathbf{x}_1, ..., \mathbf{x}_n$, so that the denominator $W(\mathbf{c})$ reaches the largest value. Hence, we obtain the following generalized statistics $W(\mathbf{c})$ for a multivariate case:

$$W_m = \left[\sum_{j=1}^{n} a_{j,n} U_{(j)} \right]^2 / (\mathbf{z}_m^T \mathbf{z}_m) \tag{12}$$

where $(\mathbf{z}_m^T \mathbf{z}_m) = \max_{1 \le j \le n} \{\mathbf{z}_j^T \mathbf{z}_j\}$ and $U_{(1)} \le ... \le U_{(n)}$, while $U_j = \mathbf{z}_m^T \mathbf{z}_j (j = 1, ..., n)$. We reject the hypothesis (1) if $W_m < W_{\alpha,n}$, where $W_{\alpha,n}$ is a quantile of the rank α of the Shapiro-Wilk distribution.

3. Generalized Shapiro-Wilk test

The generalized Shapiro-Wilk test W_p is a modification of the Shapiro-Wilk test W (1965) for a multivariate case. In the test W_p we use constants which are linear coefficients of combinations of order statistics of simple sample observations.

The construction of the test W_p proceeds as follows:

1. On the basis of the matrix $\mathbf{X}$, we create the matrix

$$\mathbf{A} = \sum_{j=1}^{n} (\mathbf{x}_j - \overline{\mathbf{x}})(\mathbf{x}_j - \overline{\mathbf{x}})^T$$

2. From among vectors $\mathbf{x}_j (j = 1, 2, ..., n)$ we select such a vector x_m, for which:

$$(\mathbf{x}_m - \overline{\mathbf{x}})\mathbf{A}^{-1}(\mathbf{x}_m - \overline{\mathbf{x}}) = \max_{1 \le j \le n} \{(\mathbf{x}_j - \overline{\mathbf{x}})^T A^{-1} (\mathbf{x}_j - \overline{\mathbf{x}})\}$$

where $\mathbf{A}^{-1}$ is a matrix inverse to the matrix $\mathbf{A}$.

3. Basing on the determined vector $\mathbf{x}_m$, we calculate:

$$U_j = (\mathbf{x}_m - \overline{\mathbf{x}})^T \mathbf{A}^{-1} (\mathbf{x}_j - \overline{\mathbf{x}}) \text{ for } j = 1, 2, ..., n$$

4. We order the value $U_{(1)} \le U_{(2)} \le ... U_{(n)}$

5. We determine a value of the statistic:

$$W_p = \frac{\left\{ \sum_{i=1}^{h} a_{(i,n)} (U_{(n-i+1)} - U_{(i)}) \right\}^2}{(\mathbf{x}_m - \overline{\mathbf{x}})^T \mathbf{A}^{-1} (\mathbf{x}_j - \overline{\mathbf{x}}))^T} \tag{13}$$

where $h=n/2$ or $h=(n-1)/2$ for even or odd n respectively, while $a_{(i,n)}(i=1,2,...,h)$ are coefficients presented in Table 1.

Table 1

Critical values for the Shapiro-Wilk multivariate normality test for $p=2$

Smoothed values			Non-smoothed values	
n	0.010	0.050	$\alpha = 0.01$	$\alpha = 0.05$
5	0.5859	0.6467	0.6002	0.6395
6	0.6232	0.6991	0.6147	0.7000
7	0.6543	0.7222	0.6667	0.7303
8	0.6808	0.7488	0.6877	0.7515
9	0.7035	0.7706	0.6950	0.7736
10	0.7232	0.7888	0.7058	0.7834
11	0.7405	0.8043	0.7163	0.7860
12	0.7558	0.8176	0.7562	0.8110
13	0.7694	0.8292	0.7592	0.8231
14	0.7816	0.8393	0.7633	0.8345
15	0.7925	0.8482	0.8043	0.8455
16	0.8025	0.8562	0.8077	0.8541
17	0.8115	0.8633	0.8081	0.8646
18	0.8198	0.8697	0.8140	0.8767
19	0.8274	0.8756	0.8205	0.8790
20	0.8344	0.8808	0.8392	0.8849

Source: own calculations.

Small values of W_p show that the distribution of a given population is not a multivariate normal distribution. Therefore, the test W_p is a left-sided test. Hence, we reject the hypothesis (5) if $W_p < W_p^\alpha$. Critical values W_p^α can be read from Tables 2 and 3.

While constructing critical values for the generalized Shapiro-Wilk test, some problems concerning smoothing these values arise (see Domański, Gadecki, Wagner, 1989).

New approximation by means of the following function was proposed:

$$W_k(x) = b_1(x-b_3)/(b_2+x).$$

It is placed in the library CurveFit in the GAUSS system.

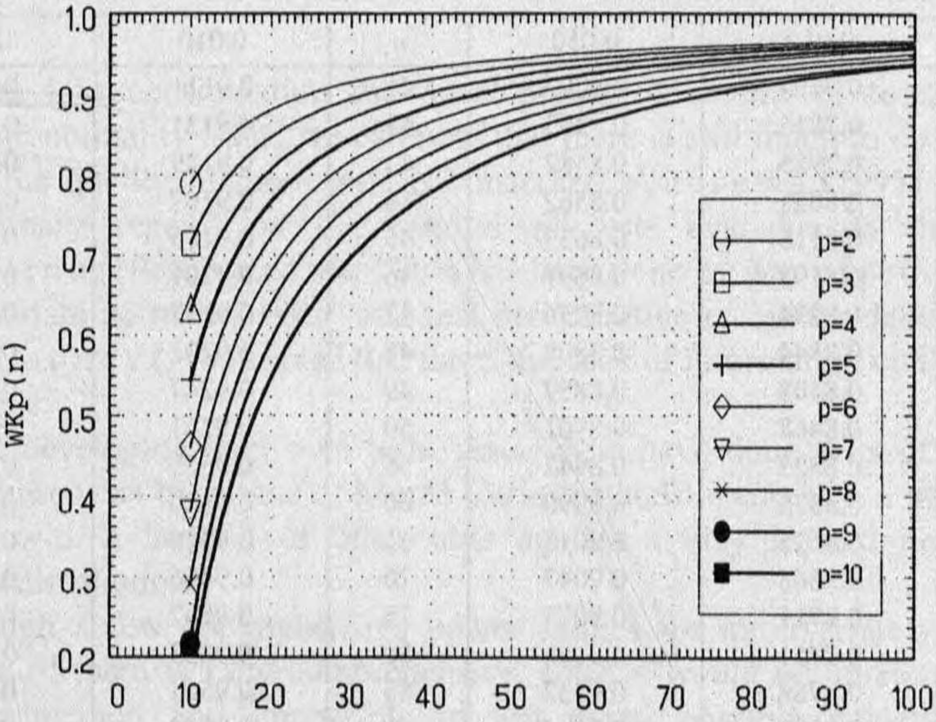

Fig 1. Curves of critical values of the Shapiro-Wilk generalized normality test for $\alpha = 0,05$ and $p = 2, 3, ..., 10$

More precise results were obtained (see Table 1 for $p = 2$), especially for small n and significance level $\alpha = 0.01$. Tables 2 and 3 present exponential critical values of the Shapiro-Wilk test of multivariate normality for two-dimensional distribution ($p = 2$) and three-dimensional distribution ($p = 3$). Figure 1 represents the curve of critical values of generalized Shapiro-Wilk test for $\alpha = 0.05$ and $p = 2, 3, ..., 10$.

Table 2

Smoothed critical values for the Shapiro-Wilk multivariate normality test for $p = 2$

n	0.010	0.050	n	0.010	0.050
5	0.5859	0.6467	33	0.8918	0.9223
6	0.6232	0.6891	34	0.8947	0.9242
7	0.6543	0.7222	35	0.8974	0.9261
8	0.6808	0.7488	36	0.8999	0.9278
9	0.7035	0.7706	37	0.9024	0.9295
10	0.7232	0.7888	38	0.9047	0.9310
11	0.7405	0.8043	39	0.9070	0.9326
12	0.7558	0.8176	40	0.9091	0.9340

Tabela 2 (contd.)

n	0.010	0.050	n	0.010	0.050
13	0.7694	0.8292	41	0.9111	0.9353
14	0.7816	0.8393	42	0.9131	0.9367
15	0.7925	0.8482	43	0.9149	0.9379
16	0.8025	0.8562	44	0.9167	0.9391
17	0.8115	0.8633	45	0.9185	0.9402
18	0.8198	0.8697	46	0.9201	0.9413
19	0.8274	0.8756	47	0.9217	0.9424
20	0.8344	0.8808	48	0.9232	0.9434
21	0.8408	0.8857	49	0.9247	0.9443
22	0.8468	0.8901	50	0.9261	0.9453
23	0.8523	0.8942	55	0.9325	0.9494
24	0.8575	0.8990	60	0.9380	0.9529
25	0.8623	0.9015	65	0.9426	0.9559
26	0.8668	0.9047	70	0.9466	0.9585
27	0.8711	0.9077	75	0.9502	0.9607
28	0.8751	0.9106	80	0.9533	0.9627
29	0.8788	0.9132	85	0.9561	0.9644
30	0.8823	0.9157	90	0.9585	0.9659
31	0.8857	0.9180	95	0.9608	0.9673
32	0.8888	0.9202	100	0.9628	0.9686

S o u r c e: own calculations.

T a b l e 3

Smoothed critical values for the Shapiro-Wilk multivariate normality test for $p = 3$

n	0.010	0.050	n	0.010	0.050
6	0.5038	0.5628	26	0.8370	0.8798
7	0.5490	0.6142	27	0.8424	0.8840
8	0.5867	0.6547	28	0.8475	0.8879
9	0.6186	0.6875	29	0.8523	0.8915
10	0.6460	0.7146	30	0.8567	0.8949
11	0.6698	0.7373	35	0.8757	0.9091
12	0.6907	0.7567	40	0.8904	0.9198
13	0.7091	0.7734	45	0.9021	0.9283
14	0.7254	0.7879	50	0.9117	0.9351
15	0.7401	0.8007	55	0.9196	0.9407
16	0.7533	0.8120	60	0.9263	0.9454
17	0.7652	0.8221	65	0.9320	0.9494
18	0.7761	0.8311	70	0.9369	0.9528
19	0.7560	0.8393	75	0.9413	0.9558
20	0.7951	0.8467	80	0.9451	0.9584
21	0.8035	0.8535	85	0.9485	0.9607
22	0.8112	0.8596	90	0.9515	0.9608
23	0.8184	0.8653	95	0.9542	0.9646
24	0.8250	0.8705	100	0.9566	0.9663
25	0.8312	0.8754			

S o u r c e: own calculations.

4. Power studies

Taking into consideration the development of dozens of tests for the multivariate normality issue, we can note that there is still much to do as far as assessing the quality of these tests is concerned. Andrews (1973) prepared a "preliminary report" on the handful of tests available at that time. D'Agostino (1986) stated that "little has been done by way of power studies for multivariate normality tests" and that no definitive recommendations could be made. Looney (1995) noted that there was lack of information on the power of these tests.

While developing their own tests, researchers have done a lot of work to compare these tests by means of Monte Carlo methods. Generally, a new test is compared with a handful of other tests against a very limited number of alternative distributions.

Although a few comprehensive power studies for multivariate normality exist, none of them is fully comprehensive, since it would be unreasonable to test every method and impossible to test every possible deviation from normality. Majority of the most comprehensive studies have deliberately limited the scope of their interest to either a particular category of tests or to considering the most popular or promising tests.

Ward (1988) compared the power of Merdia's skewness and kurtosis tests, the Malkovich-Afifi extension of the Shapiro-Wilk tests, Hawkins extension of the Anderson-Darling test, the Mardia-Foster omnibus test and two of his own proposals that extended the Kolmogorov-Smirnov and Anderson-Darling tests.

In general, Ward concluded that Mardia's skewness test, Hawkins tests and his own Anderson-Darling type test were the strongest. None of these tests, however, was good enough against the multivariate t distribution, which is a mild deviation from normality. Ward noticed that the power of the Malkovich-Afifi test, contrary to previous findings, decreased as the number of variables increased (Mardia, 1980). Ward formulated a hypothesis that the power of these procedures seemed to be related to the correlation structure, probably through the determinants of the variance-covariance matrix.

Although Mardia's tests seemed to be more effective, none of these tests was considered the best. Horswell and Looney (1992) suggested that neither affine-invariant nor coordinate-dependent tests can be regarded as superior to the others. They also questioned the "diagnostic" capabilities of this category of tests particularly effective against skewed or kurtotic alternatives. However, they stated that the performance of the skewness tests depended not only on the skewness of the distribution, but also the kurtosis. The power of skewness tests tended to be inflated when compared to alternatives with greater than normal kurtosis and depressed when compared to alternatives with less than normal kurtosis.

Romeu and Ozturk (1993) considered ten different tests of multivariate normality, the Romeu-Oztur Q_n -Cholesky and $Q_n - \Sigma^{-1}$ statistics, Mardia's tests of skewness and kurtosis, Koziol's Cramer-von Misesa test, Koziol's "radii and angles", the Cox-Small method, the Malkovich-Afifi test of skewness, Hawkins-Anderson-Darling test, and Royston's extension of the Shapiro-Wilk test. A wide range of sample sizes, dimensions, significance levels, and alternative distributions were considered.

Romeu and Ozturk compared these ten tests to sixteen different distributions, ranging from the multivariate normal to severe departures from normality. However, as always none of these tests is the best for all situations. They found that two Q_n tests and Royston's test are the best for general situations of severe or moderate non-normality. For deviations in normality due mainly to skewness, Mrdia's skewness test, Malkovich and Afifi's test and Koziol's angles test were the best. For deviations in normality due to kurtosis, Mardia's kurtosis test, Koziol's Cramer-von Mises test and Hawkins test were the best. The Cox-Small test was found to be the best for alternatives that had a mild departure from multivariate normality.

Romeu and Ozturk noted that Royston's test and especially Koziol's "radii and angles" test were sensitive to the correlation structure of the distribution and had algorithmic problems. Thus, they advised against the use of any of these tests. On the contrary, Mardia's tests were considered to be quite effective generally. The only problem with Mardia's tests was a slow convergence to the asymptotic null distribution. That is why, Romeu and Ozturk recommended (in general) the use of empirical critical values rather than asymptotic critical values. They provided a table of empirical critical values for the tests considered.

Meclin and Moundford (2003) investigated the following eight tests of multivariate normality that used asymptotic critical values:

- Mardia's test for multivariate skewness,
- Mardia's test of multivariate kurtosis,
- The Mardia-Foster C_W^2 omnibus statistic,
- The Mardia-Kent omnibus statistic,
- The Royston's multivariate Shapiro-Wilk test,
- The Romeu-Ozturk test,
- The Mudholkar-Srivastava-Lin extension of the Shapiro-Wilk test and
- The Henze-Zirkler empirical characteristic function test.

Mecklin and Mundford reasoning for preferring tests of multivariate normality that use asymptotic critical values, was that of convenience for the researcher in not having to construct one's own table of critical values.

However, Mecklin (2000) in addition to the eight tests mentioned above that used critical values, also considered the following tests that used empirical critical values:

– Hawkin's extension of the Anderson-Darling-Hawkins test,
– Koziol's extension of the Cramera-von Mises test,
– The Paulson-Roohan-Sullo version of the Anderson-Darling test,
– Singh's test of the correlation of the beta plot with classical estimates of mean and variance, and
– Singh's test of the correlation of the beta plot with robust M-estimates of mean and variance.

Mecklin and Mundfrom (2003) evaluated the power of the eight tests mentioned above in a Monte Carlo study against both the multivariate normal distribution and twenty six alternatives to normality. A wide range of sample sizes and dimensions were sampled. Some of the combinations involved sample sizes which are quite small for multivariate analysis and where asymptotic critical values may perform poorly. As it was emphasized by Romeu and Ozturk (1993) and Muldholkar and Srivastava (2002), convergence to the asymptotic distribution is often very slow for multivariate normality tests and requires n to be as large as 200. Unfortunately, Mecklin and Mundford (2003) considered only smaller values, namely $n = 20$, 51, 100. Mecklin and Mundfromk discovered that the tests of Mardia-Foster, Mardia-Kent, Romeu-Ozturk and Mudholkar-Srivastava-Lin had Type I error rates in some of the situations exceeded 0.10 (twice the nominal rate of $\alpha = 0.05$) against data generated to be multivariate normal.

5. Final remarks

The assumption that a multivariate data set stems from a multivariate normal distribution is central to many commonly employed multivariate statistical methods. If this assumption does not hold, the results of the statistical analysis become suspect. A lot of multivariate analysis are minimally acceptable, as researchers often have to use samples which are not ideal, both in terms of sample size and the methodology applied in case of these samples.

The initial attempts to test multivariate normality began over thirty years ago. Healy (1968) extended the $Q-Q$ plot to the chi-square plot often used to graphically assess multivariate normality. Mardia proposed multivariate measures of skewness and kurtosis. They are very useful both as descriptive statistics for a multivariate sample and as the basis for two very useful tests for the multivariate normality issue. Mardia's tests are probably the most popular formal procedure for goodness-of-fit to the multivariate normal distribution.

Many multivariate extensions of the standard multivariate goodness-of-fit procedures, such as the chi-square Kolmogorov-Smirnov, Cramer-von Mises and Anderson-Darling, have been proposed. Some of the most promising efforts of this type are due to Hawkins, Koziol and Paulson. Efforts have also been made to extend the Shapiro-Wilk test of multivariate normality.

The approaches to testing multivariate normality by means of either goodness-of-fit procedures or measures of skewness and kurtosis have been subject to theoretical criticism. These categories of tests were criticised for the lack of consistency against all possible alternatives and for not being "truly" multivariate procedures.

Comparisons of power for tests of multivariate normality have been carried out. However, there has been no uniformity in the tests analysed or the alternative distributions studied. The only tests that have been considered in almost every power study are the skewness and kurtosis tests of Mardia. Generally, Mardia's tests have been considered effective, although their application as a "diagnostics" in order to find a reason for non-normality was questioned by both Horswell, Looney and Henze. Other tests which are potentially useful include those of Koziol (1993), Royston (1983) and particularly Henze and Zinkler (1990).

The previous investigations revealed that none of the methods is good enough as far as multivariate normality testing is concerned. The graphic approach, such as the visual inspection of a chi-square plot or beta plot, will signal gross departures from normality and alert one to outliers. Multivariate measures of skewness and kurtosis are useful both as descriptive statistics for the multivariate data set and as the basis for normality tests. More complex procedures, such as combinations of skewness and kurtosis, generalizations of univariate goodness-of fit tests, or the newer class of consistent tests require further investigations.

Finally, it is worth drawing our attention to the asymptotic critical values, which can be also applied for very large n $(n \geq 200)$. Therefore, application of empirical critical values in the tests of multivariate normality is advisable (see section 3).

References

Andrews D. F., Gnanadesikan R. & Warner J. L. (1973), *Methods for assessing multivariate normality*, [in:] *Proceedings of the international symposium on multivariate analysis*, vol. 3, Academic Press, New York, 95–116.

Bilodeau M. & Brenner D. (1999), *Theory of multivariate statistics*, Springer, New York.

D'Agostino R. B. (1986), *Tests for the normal distribution*, [in:] *Goodness-of-Fit Techniques*, eds D'Agostino & M. A. Stephens, M. Dekker, New York, 367–419.

Domański Cz., Gadecki H., Wagner W. (1989), *Wartości krytyczne uogólnionego testu normalności Shapiro-Wilka*, „Przegląd Statystyczny", **36**, 107–112.
Domański Cz. (1990), *Testy statystyczne*, PWE, Warszawa.
Gnanadesikan R. (1977), *Methods for statistical data analysis of multivariate observations*, Wiley, New York.
Healy M. J. R. (1968), *Multivariate normal plotting*, "Appl. Statist.", **17**, 157–161.
Henze N. (2002), *Invariant tests for multivariate normality: a critical review*, "Statist. Papers", **43**, 467–506.
Henze N. & Zirkler, B. (1990), *A class of invariant consistent tests for multivariate normality*, "Comm. Statist. – Theory Methods", **19**, 3595–3618.
Horswell R. L. & Looney S. W. (1992), *Diagnostic limitations of skewness coefficients in assessing departures from univariate and multivariate normality*, "Comm. Statist. Comp. Simulation", **22**, 437–439.
Koziol J. A. (1986), *Assessing multivariate normality: a compendium*, "Comm. Statist. – Theory Methods", **15**, 2763–2783.
Koziol J. A. (1989), *A note on measures of multivariate kurtosis*, "Biom. J.", **15**, 619–624.
Koziol J. A. (1993), *Probability plots for assessing multivariate normality*, "The Statistician", **42**, 161–174.
Looney S. W. (1995), *How to use tests for univariate normality to assess multivariate normality*, "Amer. Statist.", **39**, 75–79.
Malkovich J. F. & Afifi A. A. (1973), *On tests for multivariate normality*, "J. Amer. Statist. Assoc.", **68**, 76–179.
Mardia K. V. (1980), *Tests of univariate and multivariate normality*, [in:] *Handbook of statistics*, 1, ed. P. R. Krishnaiah, Amstersam: North Holland, 297–320.
Mudholkar G. S. & Srivastava D. K. (2002), *The elusive and illusory multivariate normality*, [in:] *Advances on theoretical and methodological aspects of probability and statistics*, ed. N. Balakrishnan, Taylor & Francis, New York, 289–301.
Rao C. R. (1982), *Modele liniowe statystyki matematycznej*, PWN, Warszawa.
Romeu J. L. & Ozturk A. (1993), *A comparative study of goodness-of-fit tests for multivariate normality*, "J. Multivariate Anal.", **43**, 309–334.
Roy S. N. (1953), *On a heuristic method of test construction and its use in multivariate analysis*, "Ann. Math. Statist.", **9**, 1177–1187.
Royston J. P. (1983), *Some techniques for assessing multivariate normality based on the Shapiro-Wilk*, W. "Appl. Statist.", **32**, 121–133.
Shapiro S. S. & Wilk M. B. (1965), *An analysis of variance test for normality*, "Biometrica", **52**, 591–611.
Ward P. J. (1988), *Goodness-of-fit tests for multivariate normality*, Ph.D. Thesis, University of Alabama.

Czesław Domański

Próba oceny testów wielowymiarowej normalności

Założenie o wielowymiarowej normalności leży u podstaw klasycznej metodologii statystyki wielowymiarowej. Konsekwencje odstępstw od założenia normalności rozkładów zmiennych losowych nie są jeszcze dostatecznie poznane. Istnieje wiele różnych metod konstrukcji testów

wielowymiarowej normalności. Część tych testów stanowi rozszerzenie testów jednowymiarowej normalności.

Większość prezentowanych w literaturze przedmiotu testów wielowymiarowej normalności można podzielić na cztery kategorie:

1) procedury oparte na wykresach graficznych,
2) uogólnione testy zgodności,
3) testy oparte na miarach skośności i spłaszczenia,
4) procedury oparte na empirycznych funkcjach charakterystycznych.

W artykule będzie przedstawiona próba oceny wybranych testów zarówno z punktu widzenia ich własności, jak i możliwości ich stosowania przez badaczy nawet bez gruntownego przygotowania statystycznego.