ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 48, 1985

Czesław Domański*, Wiesław Wagner**

TESTS FOR NORMALITY BASED ON SKEWNESS AND KURTOSIS MEASURES

1. Introduction

In the theory of statistical inference a wide class of goodness of fit tests includes tests for normality. They allow to verify the goodness of fit of normal and empirical distributions of the tested random variable. The problem of the verification of normality assumptions of a distribution is of vital importance for the mathematical statistics since majority of the methods are based on this assumption.

This paper presents a class of tests for normality based on measures of the distribution shape. These measures include skewness (asymmetry) measure and kurtosis measure. On the basis of these measures the departure of the considered distribution from the normal distribution can be determined. They assume for each distribution fixed values if there are finite values of the first four central moments of the distribution. For instance, the measures of skewness in the case of symmetric distributions assume the value of zero.

* Dr., Lecturer, Institute of Econometrics and Statistics, University of Łódź.
** Dr., Lecturer, Academy of Agriculture, Poznań.

2. Measures of Asymmetry and Kurtosis

Let

(1) $$\underline{X} = (X_1, \dots, X_p)'$$

be a p-dimensional random vector with finite distribution parameters

(2) $$E\underline{X} = \underline{\mu} = (\mu_1, \dots, \mu_p)'$$
$$D\underline{X} = \underline{\Sigma} = (\sigma_{ii'}),$$

where $\underline{\Sigma}$ is a positive determined matrix.
In the case when p = 1 we shall use X, μ and σ^2, respectively.

Let

(3) $$(\underline{X}_1, \dots, \underline{X}_n) = \{\underline{X}_j\}$$

denote an n-element random sample of p-dimensional independent vectors with a uniform distribution, which are the realizations of random vector $\underline{X}$. If p = 1 the random sample is denoted as $(X_1, \dots, X_n) = \{X_j\}$. Unbiased estimators of the parameter sample $\underline{\mu}, \underline{\Sigma}$ as well as μ and σ^2 are denoted as

(4) $$\underline{\overline{X}} = (\overline{X}_1, \dots, \overline{X}_p)'$$
$$\underline{S} = (s_{ii'}) = \frac{1}{n}\sum_{j=1}^{n}(\underline{X}_j-\underline{\overline{X}})(\underline{X}_j-\underline{\overline{X}})'$$

and $\overline{X}$, S^2, respectively.

We assume that the distribution of vector $\underline{X}$ is determined by distribution function $F_p(\underline{x})$, while of variable X - distribution function F(x), where $\underline{x} \in R^p$ and $x \in R$; R^1 denotes an 1--dimensional real space. We introduce notations $\Phi_p(\underline{x})$ and $\Phi(x)$ for distribution function p-variate and univariate normal distribution. Next by $H_{op} : F_p(\underline{x}) = \Phi_p(\underline{x})$ we denote a null hypothesis stating that vector $\underline{X}$ has a p-dimensional normal distribution while for variable X we have $H_o : F(x) = \Phi(x)$.

We shall define next distribution parameters for $p = 1$

- central moment of the r-th order

(5) $$\mu_r = E[(X-\mu)^r], \qquad r = 0, 1, 2, \ldots$$

- asymmetry (skewness) coefficient

(6) $$\sqrt{\beta_1} = \mu_3/\mu_2^{3/2} \quad \text{or} \quad \beta_1 = \mu_3^2/\mu_2^3.$$

- kurtosis coefficient

(7) $$\beta_2 = \mu_4/\mu_2^2.$$

The following inequalities occur among the above mentioned coefficients

(8) $$\beta_2 > 1 + \beta_1$$
$$\beta_2 < 3 + 1.5\,\beta_1$$

The measures $\sqrt{\beta_1}$ and β_2 are applied mainly to

1) a choice of representatives in a family of distributions (e.g. in the family of Pearson's distributions),

2) a determination of tests for normality (e.g. the test based on the standardized fourth central sample moment),

3) studying the robustness of some testing procedures for departure from normal distribution (e.g. using the coefficient $\sqrt{\beta_1}$ in studying robustness of t-Student test in the verification of hypothesis $\mu = \mu_o$, where μ is an expected value in the population, and μ_o its hypothetical value).

A decisive point in introducing the distribution of t-Student statistic of a quotient form, is independence of the numerator from the denominator which occurs at the hypothesis H_o. If the sample comes from a population with non-normal distribution, then from the central limit theorem, especially from the Lindenberg--Levy theorem, it follows that the mean from the sample ($\bar{X}$) and unbiased variance estimator (S^2) has an asymptotic normal distribution (cf. [4]).

Let k_r denote the r-th cumulant in a population, where k_2 =

$= \mu_2$, $k_3 = \mu_3$, $k_4 = \mu_4 - 3\mu_3^2$. The influence of nonnormality on t statistic used in testing the hypothesis $\mu = \mu_0$ is expressed by the correlation coefficient between the variables $\bar{X}$ and S^2 of the form

$$(9) \qquad \varrho = \frac{\operatorname{cov}(\bar{X}, S^2)}{\left[D^2(\bar{X})D^2(S^2)\right]^{1/2}} = \frac{k_3/n}{\left[\frac{k_2}{n}\left(\frac{k_4}{n} + \frac{2k_2^2}{n-1}\right)\right]^{1/2}} =$$

$$= \frac{k_3}{\left[k_2\left(k_4 + \frac{2n}{n-1}k_2^2\right)\right]^{1/2}} \approx \frac{k_3}{\left[k_2\left(k_4 + 2k_2^2\right)\right]^{1/2}}$$

because at $n \to \infty$, $\frac{n}{n-1} \to 1$. If the non-normal population is symmetric, $k_3 = 0$ and hence $\varrho = 0$, then $\bar{X}$ and S^2 are asymptotically independent which allows to apply the theory of normal distribution for large n. For $k_3 \neq 0$, ϱ takes small values when k_4 is large, but $\varrho \neq 0$. Equation (9) is now written in the form

$$(10) \qquad \varrho = \frac{k_3}{\left[k_2^3\left(\frac{k_4}{k_2^2} + 2\right)\right]^{1/2}} = \frac{\sqrt{\beta_1}}{\left(\frac{k_4}{k_2^2} + 2\right)^{1/2}}$$

$$\varrho = \left(\frac{1}{2}\beta_1\right)^{1/2}$$

assuming that $k_4 = 0$.

As a result, under the above assumptions, the correlation coefficient ϱ can be treated as skewness measure. Assuming that $\sqrt{\beta_1} = 0$, we have $\varrho = 0$ and thus, the variables $\bar{X}$ and S^2 are uncorrelated. The coefficient β_2 is applied, first of all, in the verification of hypothesis that the expected value of joint variables becomes zero when there is no assumption of normality. Box and Anderson [2] using Pitman's permutation test, showed that the square of t statistic used in the verification of the above mentioned hypothesis of the expected value, has F distribution with η and $(n-1)\eta$ degrees of freedom, where

$$(11) \qquad \eta = 1 + \frac{\beta_2 - 3}{n(1-\beta_2)/n + 2} = 1 + \frac{\beta_2 - 3}{n} + o\left(\frac{1}{n}\right).$$

This result has been derived under the lack of the normality assumption for the distribution from which the sample $\{X_j\}$ was drawn.

Let us define now the basic parameters for multivariate distributions $p > 1$:

- the mixed central moment of variables $X_{i_1}, \ldots, X_{i_s}$ $(s \leq p)$ of the $(r_1 + \ldots + r_s)$-th order

$$(12) \qquad \mu^{(i_1, \ldots, i_s)}_{r_1, \ldots, r_s} = E\left[\prod_{k=1}^{s} (X_{i_k} - \mu_{i_k})^{r_k}\right],$$

where $(i_1, \ldots, i_s)$ is an arbitrary s-element subsequence from the sequence $(1, \ldots, p)$ and $r_1, \ldots, r_s = 0, 1, 2, \ldots$

- the asymmetry coefficient [11]

$$(13) \quad \beta_{1p} = \sum_{i_1, i_2, i_3 = 1}^{p} \; \sum_{i_1', i_2', i_3' = 1}^{p} \sigma^{i_1 i_1'} \sigma^{i_2 i_2'} \sigma^{i_3 i_3'} \mu_{111}^{i_1 i_2 i_3} \mu_{111}^{i_1' i_2' i_3'}$$

where $\underline{\Sigma}^{-1} = (\sigma^{ii'})$,

- kurtosis coefficient [11]

$$(14) \qquad \beta_{2p} = \sum_{i_1, i_2 = 1}^{p} \; \sum_{i_1', i_2' = 1}^{p} \sigma^{i_1 i_2} \sigma^{i_1' i_2'} \mu_{1111}^{(i_1 i_2 i_1' i_2')}$$

Besides, we introduce parameters from the sample for $p = 1$:

- the r-th order central moment

$$(15) \qquad m_r = \frac{1}{n} \sum_{j=1}^{n} (x_j - \overline{x})^r, \qquad r = 2, 3, \ldots,$$

- the asymmetry coefficient

(16) $$\sqrt{b_1} = m_3/m_2^{3/2} \quad \text{or} \quad b_1 = m_3^2/m_2^3,$$

- the kurtosis coefficient

(17) $$b_2 = m_4/m_2^2 .$$

Similarly, for $p > 1$, we have

- the mixed central moment of variables $X_{i_1}, \ldots, X_{i_s}$ of the $(r_1 + \ldots + r_s)$-th order

(18) $$m_{r_1, \ldots, r_s}^{(i_1, \ldots, i_s)} = \frac{1}{n} \sum_{j=1}^{n} \left[\prod_{k=1}^{s} (X_{i_{kj}} - \bar{X}_{i_k})^{r_k} \right],$$

- the asymmetry coefficient [11]

(19) $$b_{1,p} \sum_{i_1,i_2,i_3=1}^{p} \sum_{i_1',i_2',i_3'=1}^{p} s^{i_1 i_1'} s^{i_2 i_2'} s^{i_3 i_3'} m_{111}^{i_1 i_2 i_3} m_{111}^{i_1' i_2' i_3'},$$

where $S^{-1} = (s^{ii'})$

- the kurtosis coefficient [11]

(20) $$b_{2,p} = \sum_{i_1',i_2=1}^{p} \sum_{i_1 i_2=1}^{p} s^{i_1 i_2} s^{i_1' i_2'} m_{1111}^{(i_1 i_2 i_1' i_2')}.$$

The measures determined by formulae (19) and (20)can be given in the form of certain powers in two-linear and square forms

(21) $$b_{1,p} = \frac{1}{n^2} \sum_{j,j'=1}^{n} \left[(\underline{X}_j - \underline{\bar{X}})' \underline{S}^{-1} (\underline{X}_{j'}^{-1} - \underline{\bar{X}}) \right]^3$$

(22) $$b_{2,p} = \frac{1}{n} \sum_{j=1}^{n} \left[(\underline{X}_j - \underline{\bar{X}})' S^{-1} (\underline{X}_j - \underline{\bar{X}}) \right]^2.$$

Assuming that H_o and H_{op} are true, we have $\beta_1 = \beta_{1,p}$ and $\beta_2 =$

= 3 and $\beta_{2,p} = (p+2)p$. Hence these hypotheses can be presented in the equivalent forms:

$$H_o' : \beta_1 = 0 \wedge \beta_2 = 3 \quad \text{and} \quad H_{op} : \beta_{1,p} = 0 \wedge \beta_{2,p} = p(p+2).$$

Further on we shall construct test functions for the verification of hypotheses H_o and H_{op}.

3. Test Based on $\sqrt{b_1}$

Now, we shall discuss the attempts of determining the $\sqrt{b_1}$ distribution under the assumption of the hypothesis H_o. The best results have been obtained using Johnson's system of curves [9]. Such a result is presented by D'Agostino [5] who reduced the $\sqrt{b_1}$ statistic to a random variable with N(0,1) distribution assuming the hypothesis H_o and $n \geq 8$.

Let

$$Y = \sqrt{b_1}\left[\frac{(n+1)(n+3)}{6(n-2)}\right]^{1/2},$$

$$\beta_2(\sqrt{b_1}) = \frac{3(n^2+27n-70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)},$$

$$W^2 = -1 + \left[2\left(\beta_2(\sqrt{b_1}) - 1\right)\right]^{1/2},$$

$$\delta = 1/\left[\ln W\right]^{1/2},$$

$$\tau = \left[2(W^2-1)\right]^{1/2},$$

then the variable

$$Z = \delta \ln\left[Y/\tau + \sqrt{(Y/\tau)^2 + 1}\right] \tag{23}$$

has approximately the N(0,1) distribution.

The hypothesis H_o is rejected if $|z| > u_\alpha$, where $\Phi(u_\alpha) = 1 - \alpha/2$, and α is a given significance level.

D'Agostino and Tietjen [7] carried out

comparative studies of various approximations of the $\sqrt{b_1}$ distribution taking into account the following results (cf. Table 1):

a) their own simulation results,

b) the curves S_u (the approximation of D'Agostino [5]),

c) the approximated t-Student distribution,

d) Cornish-Fisher expression [3],

e) the modified Cornish-Fisher expression [8],

f) the approximation by normal distribution.

The approximation by t-distribution or VII-type curves from Pearson's system is as follows:

$$T = \frac{\sqrt{b_1}\left(\frac{v}{v-2}\right)^{1/2}}{\mu_2(\sqrt{b_1})}. \tag{24}$$

The statistic given in formula (24) has t-Student distribution with v degrees of freedom, with

$$v = \frac{4\beta_2\left(\sqrt{b_1}\right) - 6}{\beta_2\left(\sqrt{b_1}\right) - 3}, \quad \mu_2\left(\sqrt{b_1}\right) = 6(n-2)/n^{[3]}, \tag{25}$$

where $\sqrt{b_1}$ is determined in formula (16), $n^{[k]} = n(n-1)\ldots(n-k+1)$.

The approximation by normal distribution takes into account $\sqrt{b_1}$ as a variable with normal distribution and with its expected value equal zero and variance $\mu_2(\sqrt{b_1})$.
On the basis of Table 1, we can note that the approximation of variable $\sqrt{b_1}$ by normal distribution is of relatively small accuracy. In other cases slight differences occur in quantiles of $\sqrt{b_1}$ distribution.

The critical values for $n > 25$ were given by Pearson and Hartley [15] and for $n < 25$ by Mulholland [14] (cf. Table 2), who found them on the basis of some analytical studies on the singularity of the density function of $\sqrt{b_1}$ distribution.

D'Agostino and Tietjen [7] (cf. Table 3) also gave the critical values for n = 5(1)11,13,15,17,20,23,25,30,35, obtained using the simulation method. A comparison of the-

Table 1

Quantiles of approximated $\sqrt{b_1}$ distributions

n	Approximation	α 0.10	0.05	0.01	0.001
8	(a)	0.760	0.991	1.455	1.873
	(b)	7	-1	-34	56
	(c)	7	-1	-34	56
	(d)	8	14	-14	-17
	(e)	-4	17	19	1
	(f)	12	1	-52	-10
15	(a)	0.648	0.862	1.275	1.775
	(b)	2	-12	-13	27
	(c)	2	-12	-16	27
	(d)	0	-12	-9	48
	(e)	-1	-12	-7	49
	(f)	19	-6	-64	-167
20	(a)	0.593	0.777	1.152	1.614
	(b)	-4	-5	-2	38
	(c)	-4	-6	-4	38
	(d)	-6	-9	1	76
	(e)	-6	-9	2	76
	(f)	13	1	-52	-153
35	(a)	0.474	0.624	0.932	1.332
	(b)	1	-3	-9	-13
	(c)	1	-3	-11	-13
	(d)	0	-4	-7	-4
	(f)	12	2	-47	-156

S o u r c e: On the basis of [7].

Table 2

Quantiles of distribution of $\sqrt{b_1}$ statistic

n	α 0.05	α 0.01	n	α 0.05	α 0.01
4	0.987	1.120	15	0.851	1.272
5	1.049	1.337	16	0.834	1.247
6	1.042	1.429	17	0.817	1.222
7	1.018	1.457	18	0.801	1.199
8	0.998	1.452	19	0.786	1.176
9	0.977	1.433	20	0.772	1.155
10	0.954	1.407	21	0.758	1.134
11	0.931	1.381	22	0.746	1.114
12	0.910	1.353	23	0.733	1.096
13	0.890	1.325	24	0.722	1.078
14	0.870	1.298	25	0.710	1.060

Source: On the basis of [14].

Table 3

Quantiles of distribution of $\sqrt{b_1}$ statistic

n		5	6	7	8	9	10	11
α	0.05	1.058	1.034	1.008	0.991	0.977	0.950	0.929
	0.01	1.342	1.415	1.432	1.425	1.408	1.397	1.376

n		13	15	17	20	23	25
α	0.05	0.902	0.862	0.820	0.777	0.743	0.714
	0.01	1.312	1.275	1.188	1.152	1.119	1.073

Source: On the basis of [5].

se values with the results of Mulholland shows slight differences between them.

4. Test Based on b_2

An accurate distribution of b_2 for $n > 4$ assuming that the hypothesis H_0 is true, has not been known so far. That is why various approximations for b_2 by Johnson's S_u distribution and Pearson's IV-type distribution have been found. The approximation by S_u distribution has the following form [1]:

$$(26)\qquad Z = \begin{cases} \gamma + \delta \ln\left[\frac{b_2 - \xi}{\lambda} + \sqrt{\left(\frac{b_2 - \xi}{\lambda}\right)^2 + 1}\right], & n \geq 25 \\ \gamma + \delta \ln \frac{b_2 - \xi}{\xi + \lambda - b_2}, & n < 25 \end{cases}$$

where constants γ, δ, ξ and λ will be found using the method of moments, presented among others, by Pearson and Hartley [15]. The variable Z has approximately normal N (0,1) distribution. The verification of the hypothesis H_0 consists in a comparison of the values of Z with a corresponding value of u_α, where $\Phi(u_\alpha) = 1 - \alpha$.

Critical values of the distribution of b_2 were given by Pearson and Hartley [15] (table 34c) for $n < 200$ and $\alpha = 0.05$, 0.01. Also for the same values of α critical values were given additionally by the approximation S_u and VI-type at n = 50(25) 150, 200, 400. These values do not differ from each other up to the second place after comma. Using the simulation method D'Agostino and Tietjen [6] generated critical values for small sample sizes n = 7(1)10,12,15(5) 50 (cf. Table 4).

Table 4

Quantiles of distribution of b_2 statistic

n	α			
	0.05		0.01	
	lower	upper	lower	upper
1	2	3	4	5
7	1.41	3.55	1.25	4.23
8	1.46	3.70	1.31	4.53

Table 4 (contd.)

1	2	3	4	5
9	1.53	3.86	1.35	4.82
10	1.56	3.95	1.39	5.00
12	1.64	4.05	1.46	5.20
15	1.72	4.13	1.55	5.30
20	1.82	4.17	1.65	5.36
25	1.91	4.16	1.72	5.30
30	1.98	4.11	1.79	5.21
35	2.03	4.10	1.84	5.13
40	2.07	4.06	1.89	5.04
45	2.11	4.00	1.93	4.94
50	2.15	3.99	1.95	4.88

S o u r c e: On the basis of [6].

5. Properties of $b_{1,p}$ Statistic

Now we shall discuss the properties of the generalized skewness coefficient $b_{1,p}$:

(i) The $b_{1,p}$ statistic is invariant in relation to the orthogonal transformation $\underline{Y} = \underline{C}\,\underline{X}$. It results immediately from the form of eq. (21) to which we substitute $\underline{X}_j - \overline{\underline{X}} = \underline{C}'\,\underline{Y}_j - \overline{\underline{Y}}$.

(ii) The $b_{1,p}$ statistic is invariant in relation to the non-singular transformation $\underline{X} = \underline{A}\,\underline{Y} + \underline{b}$. It results from the form of eq. (21) and $\underline{X}_j - \overline{\underline{X}} = \underline{A}\,\underline{Y}_j - \overline{\underline{Y}}$.

(iii) The $b_{1,p}$ statistic includes $f = p(p + 1)(p + 2)/6$ distinct elements.

In the summation form of $b_{1,p}$ statistic we have 2^p elements (variation with repetitions), but only $f = \binom{p+2}{3}$ (three-element combinations with repetitions) of distinct elements.

(iv) There is an inequality

$$b_{1,p} \leqslant np^3. \tag{27}$$

Let

$$\underline{Z}_j = \underline{S}^{-1/2}(\underline{X}_j - \underline{\bar{X}}),$$

then

$$b_{1,p} = \frac{1}{n} \sum_{j,j'=1}^{n} \{\underline{Z}_j'\underline{Z}_j\}^3 \leq \frac{1}{n} \sum_{j,j'=1}^{n} p^3 = \frac{1}{n} \cdot n^2p^3 = np^3.$$

(v) The $b_{1,p}$ statistic expressed by means of angles and Mahalanobis distances assumes the form [13]

(28) $$b_{1,p} = \frac{1}{n^2} \sum_{j=1}^{n} \sum_{j'=1}^{n} (r_j r_{j'} \cos \theta_{jj'})^3,$$

where $\cos \theta_{jj'} = r_{jj'}/r_j r_{j'}$, $r_{jj'} = (r_j^2 + r_{j'}^2 - d_{jj'}^2)/2$,

$$d_{jj'} = (\underline{X}_j - \underline{X}_{j'})' S^{-1}(\underline{X}_j - \underline{X}_{j'})$$

and

$$r_j = (\underline{X}_j - \underline{\bar{X}})' S^{-1}(\underline{X}_j - \underline{\bar{X}}).$$

(vi) The expected value of $b_{1,p}$ is expressed by the formula

(29) $$E(b_{1,p}) = \frac{p(p+2)}{(n+1)(n+2)}[(n+1)(p+1) - 6].$$

This formula is given by M a r d i a [12] for $n \to \infty$, $E(b_{1,p}) \to 0$.

Due to the invariance of the linear transformation we can present the $b_{1,p}$ statistic in the form

(30) $$b_{1,p} = \sum_{i_1,i_2,i_3}^{n} \left\{m_{111}^{(i_1 i_2 i_3)}\right\}^2 =$$

$$= \left\{m_3^{(1)}\right\}^2 + \ldots + 3\left\{m_{21}^{(12)}\right\}^2 + \ldots + 6\left\{m_{111}^{(123)}\right\}^2 + \ldots$$

where:

$$m_{111}^{(i_1 i_2 i_3)} = m_{12}^{(i_1 i_2)}, \quad i_1 \neq i_2,$$

$$m_{111}^{(i_1 i_1 i_1)} = m_3^{(i_1)}$$

while

$$m_3^{(1)} = \frac{1}{n} \sum_{j=1}^{n} (x_{1i} - \bar{x}_1)^3,$$

$$m_{21}^{(12)} = \frac{1}{n} \sum_{j=1}^{n} (x_{1j} - \bar{x}_1)(x_{2j} - \bar{x}_2),$$

$$m_{111}^{(123)} = \frac{1}{n} \sum_{j=1}^{n} (x_{1j} - \bar{x}_1)(x_{2j} - \bar{x}_2)(x_{3j} - \bar{x}_3).$$

Assuming that the hypothesis that the sample $\{\underline{X}_j\}$ comes from a multidimensional normal population $N_p(\underline{0}, \underline{I})$, is true, we have moments up to the n^{-1}-th order, of the form [11]

(31) $$E\left(m_{111}^{i_1 i_2 i_3}\right) = 0,$$

$$D^2\left(m_3^{(1)}\right) = 6/n,$$

$$D^2\left(m_{21}^{(12)}\right) = 2/n,$$

$$D^2\left(m_{111}^{(123)}\right) = 1/n,$$

$$\operatorname{cov}\left(m_{111}^{(i_1 i_2 i_3)}, m_{111}^{(i_1' i_2' i_3')}\right) = 0, \quad i_1, i_2, i_3 \neq i_1', i_2', i_3'.$$

Note, that the assumption of $\underline{\mu} = \underline{0}$ and $\underline{\Sigma} = \underline{I}$ is possible due to the property (ii).

Let $f = p(p+1)(p+2)/6$-dimensional vector be given

(32) $\underline{M} = \left(m_3^{(1)}, \ldots, \sqrt{3}\, m_{21}^{(12)}, \ldots, \sqrt{6}\, m_{111}^{(123)}, \ldots \right)'$

then

(33) $$b_{1,p} = \underline{M}'\underline{M}$$

On the basis of formulae (31) we have

(34) $$E(\underline{M}) = 0$$

$$D(\underline{M}) = E(\underline{M}\underline{M}') = \text{diag}\,(6/n, \ldots, 6/n, \ldots, 6/n) = (6/n)\underline{I}.$$

Hence

(35) $$\underline{M} \sim N_f(\underline{0}, 6/n\ \underline{I}),$$

while

(36) $$n\underline{M}'\underline{M}/6 = nb_{1,p}/6 \sim \chi^2_f.$$

Formulae (35) and (36) occur when $n \to \infty$. An accurate distribution of the variable $b_{1,p}$ is not yet known. Besides no other approximations of the variable $b_{1,p}$ are known as in the univariate case. For $p > 7$ the following approximation can be applied

(37) $$(2nb_{1,p}/6)^{1/2} \sim N(2f-1, 1).$$

M a r d i a [12] determined the critical values for the distribution of $b_{1,p}$ using the Monte-Carlo method for n = 10(2)20 (5), 30(10), 100(50), 200(100), 400(200), 1000(500), 3000 (1000), 5000 and α = 0.001, 0.01, 0.025, 0.05, 0.075, 0.10 (cf. Table 5). For p = 3 and p = 4 Mardia determined the critical values, however they have not been published.

Table 5

Quantiles of distribution of $b_{1,2}$ statistic

n		10	12	14	16	18	20	25	30	40	50
α	0.05	3.694	3.319	3.031	2.775	2.556	2.356	1.969	1.687	1.319	1.069
	0.01	5.938	4.938	4.581	4.231	3.962	3.669	3.106	2.681	2.087	1.744

Source: On the basis of [12].

Table 6

Quantiles of distribution of $b_{2,2}$ statistic

n		10	12	14	16	18	20	25	30	40	50
α	0.05	4.887 9.203	5.053 9.593	5.179 9.769	5.318 9.941	5.382 10.005	5.533 10.114	5.689 10.159	5.855 10.156	6.139 10.109	6.239 9.987
	0.01	4.580 10.378	4.732 10.881	4.842 11.159	4.977 11.387	5.045 11.478	5.175 11.609	5.351 11.628	5.518 11.594	5.703 11.453	5.909 11.181

Source: On the basis of [12].

6. Properties of the $b_{2,p}$ Statistic

For the $b_{2,p}$ statistic the following properties occur.

(i) The $b_{2,p}$ statistic is invariant due to the orthogonal transformation $\underline{Y} = \underline{C}\,\underline{X}$ and non-singular $\underline{X} = \underline{A}\,\underline{Y} + \underline{b}$.

(ii) The expected value of $b_{2,p}$ assumes the form [11]

(38) $$E(b_{2,p}) = \frac{p(p+2)(n-1)}{n+1}.$$

(iii) The variance of $b_{2,p}$ is determined by the formula [12]

(39) $$D^2(b_{2,p}) = \begin{cases} \dfrac{8p(p+2)(n-3)}{(n+1)^2(n+3)(n+5)}\,(n-p-1)(n-p+1) \\ \dfrac{8p(p+2)}{n}, \text{ at } n^{-1}. \end{cases}$$

The first formula was introduced by taking into account the multivariate beta distribution, and the second one by using L a w l e y's method [10].

(iv) $b_{2,p}$ can be expressed in the form

(40) $$b_{2,p} = \frac{1}{n}\sum_{j=1}^{n} r_j^4,$$

where r_j is Mahalanobis distance between $\underline{X}_j$ and $\underline{X}$.

Taking formulae (38) and (39) we can obtain two tests verifying the hypothesis H_{op}, whose statistics are as follows.

(41) $$N_1 = \frac{\{(n+1)b_{2,p} - p(p+2)(n-1)\}\cdot\{(n+3)(n+5)\}^{1/2}}{\{8p(p+2)(n-3)(n-p-1)(n-p+1)\}^{1/2}}$$

for the accurate variance $D^2(b_{2,p})$ and

(42) $$N_2 = \frac{b_{2,p} - p(p+2)}{\{8p(p+2)/n\}^{1/2}}$$

for the approximated variance $D^2(b_{2,p})$ up to the n^{-1} order.

Statistics (41) and (42) have the N(0,1) distribution by virtue of the central limit theorem.

An accurate distribution of the variable $b_{2,p}$ under the assumption that the hypothesis H_{op} is true, is unknown. The necessary critical values for the distribution of $b_{2,2}$ have been generated by M a r d i a [12] using the Monte-Carlo method in the same range of n as for $b_{1,2}$, and $\alpha = 0.01$, 0.025, 0.05, 0.10 giving two values - upper and lower. Table 6 presents these values for $n < 50$.

BIBLIOGRAPHY

[1] B o w m a n K. O., S h e n t o n L. R., (1975), Omnibus Test Contours for Departures from Normality Based on $\sqrt{b_1}$ and b_2, "Biometrika" 62, p. 243-250.

[2] B o x G. E. P., A n d e r s o n S. L. (1955), Permutation Theory in the Derivation of Robust Criteria and the Study of Departures from Assumptions, J.R.S.S.B. 17, p. 1-34.

[3] C o r n i s h E. A., F i s h e r R. A. (1937), Moments and Cumulants in the Specification of Distributions, Rev. Inst. Int. Statist. 5, p. 307.

[4] C r a m e r H. (1958), Metody matematyczne w statystyce, PWN, Warszawa.

[5] D'A g o s t i n o R. B., T i e t j e n G. L. (1973), Approaches to the Null Distribution of $\sqrt{b_1}$, "Biometrika" 60, p. 169-175.

[6] D'A g o s t i n o R. B., T i e t j e n G. L. (1971), Simulation Probability Points of b_2 for Small Samples, "Biometrika" 58, p. 669-672.

[7] D'A g o s t i n o R. B. (1970), Transformation to Normality of the Null Distribution of g_1, "Biometrika" 57, p. 679--681.

[8] G e a r y R. C. (1947), The Frequency Distribution of $\sqrt{b_1}$ for Samples of all Sizes Drawn at Random from a Normal Population, "Biometrika" 34, p. 68-97.

[9] J o h n s o n N. L. (1949), Systems of Frequency Curves Generated by Methods of Translation, "Biometrika" 36, p. 149--176.

[10] L a w l e y D. N. (1959), Test of Significance in Canonical Analysis, "Biometrika" 46, p. 59-66.

[11] M a r d i a K. V. (1970), Measures of Multivariate Skewness and Kurtosis with Applications, "Biometrika" 57, p. 519-530.

[12] M a r d i a K. V. (1974), Applications of Some Measures of Multivariate Skewness and Kurtosis in Testing Normality and Robustness Studies, "Sankhya" 36, ser. B., p. 115-128.

[13] M a r d i a K. V. (1975), Assessment of Multinormality and the Robustness of Hotelling's T^2 Test, Appl. Statist. 24, p. 163-171.

[14] M u l h o l l a n d H. P. (1977), On the Null Distribution of $\sqrt{b_1}$ for Samples of Size at Most 25, with Tables, "Biometrika" 64, p. 401-409.

[15] P e a r s o n E. S., H a r t l e y H. O. (1966), Biometrika Tables for Statisticians, Vol. 1, Cambridge University Press.

[16] P e a r s o n E. S., H a r t l e y H. O. (1972), Biometrika Tables for Statisticians, Vol. 2, Cambridge University Press.

Czesław Domański, Wiesław Wagner

TESTY NORMALNOŚCI OPARTE NA MIARACH SKOŚNOŚCI I SPŁASZCZENIA

W artykule przedstawiono testy weryfikujące hipotezę o normalności rozkładu zarówno jednowymiarowego, jak i wielowymiarowego, oparte na miarach skośności i spłaszczenia. Do większości omawianych testów podano niektóre kwantyle rozkładów funkcji testowych. Zamieszczono również podstawowe własności uogólnionej miary skośności - $b_{1,p}$ oraz miary kurtozy - $b_{2,p}$.