Colloquium Biometricum 40
2010, 95–103

SIMULATION STUDY ON MULTIVARIATE NORMALITY BASED ON SHAPIRO-WILK STATISTIC

Zofia Hanusz, Joanna Tarasińska

Department of Applied Mathematics and Computer Science
University of Live Sciences in Lublin
Akademicka 13, 20-950 Lublin
e-mails: zofia.hanusz@up.lublin.pl, joanna.tarasinska@up.lublin.pl

Summary

The paper concerns three tests for multivariate normality based on the Shapiro-Wilk W statistic for the principal components of a covariance matrix. Two of them were proposed by Srivastava and Hui (1987), the third was introduced by Hanusz and Tarasińska (2008b). The type I errors of these tests at significance levels 0.1, 0.05 and 0.01 are evaluated both for the sample and residuals in the two data groups. The powers of the tests under consideration against chosen alternative distributions are also presented in both the sample and residual cases.

Keywords and phrases: test for multivariate normality, Shapiro-Wilk *W* statistic, Type I error, power of the test, Srivastava and Hui tests

Classification AMS 2010: 62H15

1. Introduction

The Shapiro-Wilk W statistic (1965)

$$W = \frac{\left[\sum_{j=1}^{\left[\frac{n}{2}\right]} a_{j,n} \left(x_{(n+1-j)} - x_{(j)} \right) \right]^2}{\sum_{j=1}^{n} \left(x_j - \overline{x} \right)^2}$$

is considered by many authors as the best statistic for checking univariate normality of data, especially for small sample sizes. In this formula $x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$ are ordered values of the sample $x_1, x_2, \ldots, x_n$. Small values of *W* indicate nonnormality. The constants $a_{j,n}$ and critical values $W_{\alpha,n}$ for $n = 3(1)50$ are tabulated in Shapiro and Wilk (1965). Royston (1982) introduced approximation for $a_{j,n}$ in the case $n > 50$. Such tables together with critical values $W_{\alpha,n}$ and many other results for the Shapiro-Wilk test can be found, among other places in Wagner (1990).

Shapiro and Wilk also proposed another test based on the following transformation of the *W* statistic, using Johnson's (1949) S_B distribution

$$G(W) = \gamma + \delta \ln\left(\frac{W - \varepsilon}{1 - W} \right) , \tag{1.1}$$

where ln denotes a natural logarithm and $G(W)$ is approximately distributed as standard normal. Tables with γ, δ and ε for sample sizes $n = 3(1)50$ are given in Shapiro and Wilk (1968) as well as in many other papers, for example in Wagner and Błażczak (1992). If $n > 50$ then estimated values of γ, δ and ε can be obtained through the outcomes of Shapiro and Francia (1972) or Royston (1982). The lower tail of normal distribution indicates nonnormality.

Statistic (1.1) was adopted by Srivastava and Hui (1987) for the multivariate case, who introduced two tests as follows.

Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ be $(p \times 1)$ independent random vectors with an unknown expected value of $\boldsymbol{\mu}$ and a covariance matrix of $\boldsymbol{\Sigma}$. Let $\overline{\mathbf{x}} = \frac{1}{n} \sum_{j=1}^{n} \mathbf{x}_j$ and

$\mathbf{S} = \frac{1}{n}\sum_{j=1}^{n}(\mathbf{x}_j - \overline{\mathbf{x}})(\mathbf{x}_j - \overline{\mathbf{x}})'$ be the sample mean and sample covariance matrix, respectively. Define $y_{ij} = \mathbf{h}_i'\mathbf{x}_j$, $i = 1,\ldots,p$; $j = 1,\ldots,n$, where $\mathbf{h}_i$ are the eigenvectors of **S**, corresponding to eigenvalues u_i. Now let us take *p* univariate Shapiro-Wilk statistic

$$W(i) = \frac{1}{nu_i}\left[\sum_{j=1}^{\left[\frac{n}{2}\right]} a_{j,n}\left(y_{i(n+1-j)} - y_{i(j)}\right)\right]^2 \quad i = 1,\cdots,p,$$

where $y_{i(1)} \leq \ldots \leq y_{i(n)}$ are ordered statistics for the i-th principal component. The first test statistic for multivariate normality proposed by Srivastava and Hui (1987) is the following:

$$M_1 = -2\sum_{i=1}^{p}\ln[\Phi(G_i)], \tag{1.2}$$

where $G_i = G(W(i))$ is the transformation (1.1) and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. Under normality, M_1 is approximately χ^2_{2p} distributed. Large values of M_1 indicate nonnormality. It should be emphasized that the maximum likelihood estimate of **S** should be applied in the calculation, not the unbiased estimate. This fact was pointed out by Hanusz and Tarasińska (2008a).

The other test for multivariate normality proposed by Srivastava and Hui (1987) is based on the statistic

$$M_2 = \min_{i=1,\ldots,p}\{W(i)\}. \tag{1.3}$$

The distribution of M_2, under normality, is approximately given by

$$\Pr(M_2 \leq x) = 1 - [1 - \Phi(G(x))]^p. \tag{1.4}$$

This test rejects normality for small values of M_2.

Hanusz and Tarasińska (2008b) introduced another test statistic for testing multivariate normality, in accordance with Srivastava and Hui's (1987, 2002) idea of Shapiro-Wilk statistics for the principal components of covariance matrix **S** with the use of transformation (1.1). This test has a simplier form than (1.2), namely $V = \sqrt{p}\,\overline{G}$, where $\overline{G} = \frac{1}{p}\sum_{i=1}^{p} G_i$. Under normality, the statistic V displays asymptotic standard normal distribution. Some preliminary results concerning this test were given in Hanusz and Tarasińska (2009). Namely, type I error at a significance level of 0.05 for V and powers of M_1, M_2 and V were evaluated against chosen alternative distributions through simulation studies based on 1000 generated samples.

The present paper is a continuation of those investigations. Now, type I errors for all three tests M_1, M_2 and V at significance levels 0.1, 0.05 and 0.01 are evaluated for both samples and residuals in two groups of data with equal and unequal numbers of observations. The powers for the three tests under consideration against chosen alternatives are also determined both in the case of samples and residuals.

All simulations were conducted in the R program (R Development Core Team, 2008) with 10,000 data sets being generated in each case.

2. Type I error study

In order to evaluate type I errors in the sample case for test statistics M_1, M_2 and V 10,000 random samples of size $n = 10$ and $n = 20$ from a p-variate normal distribution ($p = 2$ and $p = 3$) were generated. Nominal significance levels were taken as 0.1, 0.05 and 0.01. The I type errors were determined as the fraction of samples for which, respectively,

— the values of M_1 exceeded $(1-\alpha)$-th quantile of χ^2_{2p} distribution,

— the values of M_2 were less than $\frac{\varepsilon + c(p,\alpha)}{1 + c(p,\alpha)}$, $c(p,\alpha) = \exp\left(\frac{\Phi^{-1}\left(1-\sqrt[p]{1-\alpha}\right)-\gamma}{\delta}\right)$, $\gamma, \delta, \varepsilon$ are constants from

(1.1), $\Phi(\cdot)$ is the cumulative distribution function of standard normal distribution,

— the values of V were less than $\Phi^{-1}(\alpha)$, i.e. α-th quantile of standard normal distribution.

Next, the case of two groups of data was considered, for which 10,000 data sets were generated according to the following linear model:

$$\mathbf{X} = \mathbf{AB} + \mathbf{E}, \tag{2.1}$$

where $\mathbf{X}$ is a $n \times p$ matrix of the data, $\mathbf{A} = \begin{bmatrix} \mathbf{1}_{n_1}, & \mathbf{0}_{n_1} \\ \mathbf{0}_{n_2}, & \mathbf{1}_{n_2} \end{bmatrix}$ is a $n \times 2$ matrix, $n_1 + n_2 = n$, $\mathbf{B}' = [\mathbf{0}_p, \mathbf{1}_p]$, $\mathbf{E} \sim N(\mathbf{0}, \mathbf{I}_n \otimes \mathbf{\Sigma})$, $\mathbf{1}_k$ and $\mathbf{0}_k$ are the vectors of k ones and k zeros, respectively, $\mathbf{\Sigma}$ is a $p \times p$ covariance matrix, $\mathbf{I}_n$ is the identity matrix and $\otimes$ denotes the Kronecker product of the matrices. Thus n_1 observations had $N(\mathbf{0}_p, \mathbf{\Sigma})$ distribution and n_2 observations had $N(\mathbf{1}_p, \mathbf{\Sigma})$ distribution. As the distributions of all test statistics do not depend on the covariance matrix, the $\mathbf{\Sigma} = \mathbf{I}_p$ was taken in simulation. The type I errors for residuals $\left[\mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\right]\mathbf{X}$ were evaluated in the same way as in the case of the samples. Table 1 presents the results of the simulation, both for samples and residuals.

Firstly let us notice that there are no great differences between type I errors for residuals in the cases of equal and unequal numbers of observations. Statistic V turns out to be particularly robust in this case. No such difference for V is significant at a level of 0.05 (the test for two fractions was applied to conclude this).

Next, let us notice that in the case of $n = 10$ the type I errors for residuals (equal numbers of observations) are smaller than type I errors for the sample. Most of these differences are significant at level 0.05. This does not pertain to the case of $n = 20$.

Table 1. Type I errors in testing multivariate normality at significance level α

			Sample					Residuals		
			α					α		
n	p	test	0.1	0.05	0.01	n_1	n_2	0.1	0.05	0.01
10	2	M_1	0.0969	0.0400	0.0084	5	5	0.0833	0.0381	0.0057
		M_2	0.0922	0.0427	0.0089			0.075	0.0361	0.0042
		V	0.1032	0.0489	0.0082			0.0905	0.0427	0.0073
		M_1				2	8	0.0824	0.0408	0.0085
		M_2						0.0782	0.0394	0.0088
		V						0.0901	0.0425	0.0079
	3	M_1	0.0994	0.0445	0.0091	5	5	0.0782	0.0357	0.0047
		M_2	0.0891	0.0438	0.0089			0.0755	0.0302	0.0032
		V	0.1047	0.0475	0.0081			0.0893	0.0412	0.0071
		M_1				2	8	0.0868	0.0427	0.0083
		M_2						0.0800	0.0408	0.0075
		V						0.0908	0.0429	0.0079
20	2	M_1	0.1004	0.0478	0.0114	10	10	0.1019	0.0508	0.0088
		M_2	0.0958	0.0480	0.0111			0.1023	0.0499	0.009
		V	0.1017	0.0480	0.0103			0.1057	0.0480	0.0094
		M_1				5	15	0.1015	0.0499	0.0097
		M_2						0.1026	0.0491	0.0096
		V						0.1018	0.0516	0.0091
	3	M_1	0.1017	0.0477	0.0099	10	10	0.1006	0.0455	0.0105
		M_2	0.0979	0.0463	0.0121			0.0971	0.0447	0.0106
		V	0.1035	0.0487	0.0102			0.102	0.0459	0.0097
		M_1				5	15	0.1021	0.0484	0.0108
		M_2						0.0995	0.0465	0.0106
		V						0.1017	0.0509	0.0109

3. Power study

To compare the power of considered tests M_1, M_2 and V, 10,000 data sets of sizes $n = 10$, $n = 20$ and $n = 40$ from selected p-variate ($p = 2$, $p = 3$) distributions were generated for samples and the linear model (2.1). Powers were determined as the fractions of samples or residuals for which the test statistics fall into the critical areas described in Section 2. The significance level 0.05 was only considered. The following distributions for sample and error **E** in the linear model (2.1) were taken: uniform on the p-th sphere (MPII i.e. Pearson Type II), multivariate t distribution with 2 degrees of freedom (MPVII i.e. Pearson Type VII) and distribution with independent marginals of chi-square distri-

Table 2. Powers evaluated on basis of 10,000 data sets generated according to the uniform on the p-sphere, multivariate t_2 and $(\chi_3^2)^p$ distributions

			Sample					Residuals		
n	p		uniform	mult. t_2	$(\chi_3^2)^p$	n_1	n_2	uniform	mult. t_2	$(\chi_3^2)^p$
10	2	M_1	0.060	0.385	0.308	5	5	0.041	0.264	0.175
		M_2	0.045	0.381	0.276			0.033	0.264	0.162
		V	0.073	0.349	0.296			0.050	0.226	0.167
		M1				2	8	0.041	0.308	0.192
		M2						0.031	0.305	0.182
		V						0.050	0.274	0.184
	3	M1	0.049	0.439	0.279	5	5	0.038	0.284	0.145
		M2	0.035	0.435	0.252			0.028	0.295	0.130
		V	0.066	0.379	0.253			0.049	0.226	0.143
		M1				2	8	0.040	0.353	0.185
		M2						0.029	0.355	0.173
		V						0.052	0.295	0.167
20	2	M1	0.115	0.682	0.654	10	10	0.084	0.636	0.557
		M2	0.086	0.675	0.615			0.066	0.634	0.517
		V	0.134	0.642	0.638			0.099	0.592	0.539
		M1				5	15	0.085	0.635	0.558
		M2						0.065	0.627	0.527
		V						0,100	0,597	0,541
	3	M1	0,081	0,759	0,633	10	10	0,064	0,706	0,515
		M2	0,048	0,751	0,576			0,042	0,698	0,474
		V	0,106	0,701	0,578			0,081	0,640	0,478
		M1				5	15	0,067	0,717	0,531
		M2						0,045	0,705	0,489
		V						0,090	0,650	0,488
40	2	M1	0,245	0,901	0,871	20	20	0,180	0,882	0,844
		M2	0,193	0,896	0,852			0,142	0,878	0,821
		V	0,268	0,879	0,857			0,200	0,857	0,825
		M1				5	35	0.192	0.884	0.852
		M2						0.150	0.877	0.829
		V						0.213	0.857	0.836
	3	M1	0.113	0.948	0.855	20	20	0.092	0.935	0.826
		M2	0.081	0.941	0.828			0.061	0.930	0.790
		V	0.135	0.923	0.815			0.108	0.902	0.782
		M1				5	35	0.096	0.938	0.829
		M2						0.066	0.933	0.800
		V						0.114	0.905	0.787

bution with three degrees of freedom. Thus fat-tailed, light-tailed and skewed distributions were considered. MPVII and MPII distributions were generated according to Johnson (1987). The results are given in Table 2 and we may conclude as follows.

1. For multivariate t_2 and $(\chi_3^2)^p$ distributions M_1 has the highest power in most cases whereas the test based on the statistic *V* is the best for uniform on the *p*-th sphere distribution. It is worth noting that all tests rather seldom detect nonnormality when the uniform distribution is true.

2 All tests are less powerful when applied to residuals than in the case of the sample.

3 If we consider the influence of *p* on the power it may be said that the power of all tests is higher for p=2 than for p=3, both for sample and residuals for uniform and $(\chi_3^2)^p$. The contrary conclusion may be drawn for multivariate t_2 distribution.

4 If we consider the influence of equal and unequal numbers of observations in the groups on the power of the tests applied to residuals then we can note that for $(\chi_3^2)^p$ distribution the powers are higher when the groups are unequal. The same is true for t_2 when $n = 10$ and $n = 20$ (only for $p = 3$) and for the uniform distribution when n=40.

References

Hanusz Z., Tarasińska J. (2008 a). A note on Srivastava and Hui's tests of multivariate normality. *Journal of Multivariate Analysis* 99, 2364-2367.

Hanusz Z., Tarasińska J. (2008 b). Remarks on approximated tests based on Shapiro-Wilk's statistic. *Colloquium Biometricum* 38, 87-93.

Hanusz Z., Tarasińska J. (2009). Simulation study for a test of multivariate normality based on Shapiro-Wilk's statistic. *Colloquium Biometricum* 39, 45-51.

Johnson M.E. (1987). *Multivariate statistical simulation*. J. Wiley & Sons Inc, New York.

Johnson N.L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika* 36, 149-176.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Royston J.P. (1982). An extension of Shapiro and Wilk's W test for normality to large samples. *Appl. Statist*. 31, 115-124.

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

Shapiro S.S., Francia R.S. (1972). An approximate analysis of variance test for normality. JASA 67, 215-216.

Srivastava M.S. (2002). *Methods of multivariate statistics*. J. Wiley & Sons, New York.

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

Wagner W. (1990). *Test normalności wielowymiarowej Shapiro-Wilka i jego zastosowania w doświadczalnictwie rolniczym*. Wydawnictwo AR, Poznań.

Wagner W., Błażczak P. (1992). *Statystyka matematyczna z elementami doświadczalnictwa*. Wydawnictwo AR, Poznań.

BADANIA SYMULACYJNE WIELOWYMIAROWEJ NORMALNOŚCI OPARTE NA STATYSTYCE SHAPIRO-WILKA

Streszczenie

W pracy rozważa się trzy testy wielowymiarowej normalności oparte na statystyce Shapiro-Wilka dla składowych głównych macierzy kowariancji. Dwa z nich zaproponowali Srivastava i Hui (1987), a trzeci Hanusz i Tarasińska (2008b). Ocenia się błędy I rodzaju dla tych testów na poziomach istotności 0,05; 0,1 i 0,01 zarówno w przypadku próby losowej jak i reszt z modelu liniowego dla dwóch grup danych. Podane są także moce tych testów przy wybranych rozkładach alternatywnych w przypadku próby i reszt.

Słowa kluczowe: testy wielowymiarowej normalności, statystyka Shapiro-Wilka, błąd I rodzaju, moc testu, testy Srivastavy i Hui.

Klasyfikacja AMS 2010: 62H15