A Powerful Test for Multivariate Normality

Ming Zhou[a] and **Yongzhao Shao**[b,*]

[a]Department of Statistics, Iowa State University, Ames, Iowa, USA

[b]Division of Biostatistics, New York University School of Medicine, New York, USA

Abstract

This paper investigates a new test for normality that is easy for biomedical researchers to understand and easy to implement in all dimensions. In terms of power comparison against a broad range of alternatives, the new test outperforms the best known competitors in the literature as demonstrated by simulation results. In addition, the proposed test is illustrated using data from real biomedical studies.

Keywords

Goodness of fit; Normal distribution; Projection; Shapiro-Wilk test; Power

1. Introduction

Normal distribution is widely used in many applications. The problem of testing whether a sample of observations comes from a normal distribution has been studied extensively by many generations of statisticians, including [3, 6, 10, 13, 16, 17, 20, 23]. For instance, in a recent monograph devoted to the topic of testing for normality, Thode [23] reviewed more than thirty formal procedures proposed specifically for testing normality. Briey, in terms of power performance against a broad range of alternatives, the Shapiro-Wilk (SW) test [20] is the benchmark of omnibus tests for univariate data [3, 6, 23]. For testing multivariate normality, the Henze-Zirkler (HZ) test [13] is recommended by Thode [23, pp. 220]. In many practical applications, researchers often prefer to use tests that are both informative and easy to understand [4]. Although generally quite powerful as a multivariate test, the HZ test has the drawback of not being as easy to understand as the simple skewness or kurtosis based tests, or as the SW test which is known to many researchers and is generally powerful to detect outliers or influential observations as well as to skewed distributions. To the best of our knowledge, there is no known test that is both informative and has competitive power in all dimensions. In this paper, we introduce a simple informative test that is easy to understand and to implement in all dimensions. Simulations studies indicate that the new test has very competitive power compared to the HZ test and other best known tests in both univariate and multivariate cases.

The rest of this paper is organized as follows. Section 2 gives a brief review of some well-known tests for normality. Section 3 introduces our new test. Numerical studies on power comparison against a broad range of alternatives are reported in Section 4. Real data examples are given to illustrate the newly proposed test in Section 5. Some concluding remarks are in Section 6.

*Corresponding author at: Division of Biostatistics, New York University SOM, 650 First Avenue, NY 10016, USA. shaoy01@nyu.edu.

2. Some Well-known Tests for Normality

Consider independent observations $X_1,\ldots,X_n$ from a p-variate random vector X, we want to test H_0 : X has a p-variate normal density, against the general alternative H_1 : X has a p-variate non-normal Lebesgue density. Throughout this paper, we consider the cases where $n > p$. Moreover, if we say that H_0 is rejected when the test statistic is too extreme, we mean that, towards the direction of opposing normality, it exceeds the critical value, which can be accurately obtained by simple Monte Carlo simulations. Let $\overline{X}_n$ and S_n be the sample mean and sample covariance, respectively. Denote the transpose of any vector x by x^{T}, and its norm by $\|x\| = (x^{\mathrm{T}}x)^{1/2}$. Given the existence of a p-variate Lebesgue density of X, S_n is nonsingular, i.e. S_n^{-1} exists, almost surely [8]. Thus, without loss of generality, we can reject the null hypothesis of normality when S_n is singular. In the rest of this paper we consider the case where S_n is nonsingular, use $S_n^{-1/2}$ to denote the symmetric square root of S_n^{-1}, and consider the standardized data:

$$Y_i = S_n^{-1/2}(X_i - \overline{X}_n), i=1,\ldots,n. \quad (1)$$

2.1 The Shapiro-Wilk test

The Shapiro-Wilk (SW) test [20] was originally designed for testing univariate normality. Given univariate data $z_1,\ldots,z_n$, the SW statistic has the following form:

$$W_n(z_1,\ldots,z_n) = \frac{\left\{\sum_{i=1}^n a_{n,i} z_{(i)}\right\}^2}{\sum_{i=1}^n \left(z_i - \overline{z}_n\right)^2}, \quad (2)$$

where $z_{(1)},\ldots,z_{(n)}$ are the order statistics of $z_1,\ldots,z_n$, $\overline{z}_n$ is the sample mean, and the constants $a_{n,i}$ are $(a_{n,1},\ldots,a_{n,n}) = (m^{\mathrm{T}}V^{-1}V^{-1}m)^{-1/2}m^{\mathrm{T}}V^{-1}$, with $m = (m_1,\ldots,m_n)^{\mathrm{T}}$ and V the mean and covariance of the order statistics of a random standard normal sample of size n, respectively. The SW statistic can be easily evaluated using the free software R and many other statistical packages.

The SW test statistic can be regarded as the ratio of two variance estimators, the best linear unbiased estimator (BLUE) and the maximum likelihood estimator (MLE). Specifically, if $X_1,\ldots,X_n$ are independent and identically distributed (i.i.d) as $N(\mu,\sigma^2)$, then $X_i = \mu + \sigma Z_i$, $i = 1,\ldots,n$, where Z_i's are i.i.d $N(0, 1)$. The order statistics then have similar identities, namely $X_{(i)} = \mu + \sigma Z_{(i)}$. Thus, $X_{(i)} = \mu + \sigma m_i + \varepsilon_i$, where $m_i = E(Z_{(i)})$ and $\varepsilon_i = \sigma(Z_{(i)} - m_i)$. Under the normality assumption, the joint distribution of $(\varepsilon_1,\ldots,\varepsilon_n)^{\mathrm{T}}$ has zero mean and covariance matrix $\sigma^2 V_n$, where V_n is the covariance matrix of the order statistics $(Z_{(1)},\ldots,Z_{(n)})^{\mathrm{T}}$ from $N(0, 1)$. In addition, under the normality assumption, one can estimate the unknown parameter σ by using either the MLE s or the BLUE $\hat{\sigma}_{BL}$. Then W_n in (2) has an equivalent form as the variance ratio $\hat{\sigma}_{BL}^2 / s^2$. Under the alternative model of non-normality, the BLUE $\hat{\sigma}_{BL}$ tends to be smaller than the MLE s. Thus, the SW test rejects the hypothesis of normality for small values of W_n.

Moreover, from the derivation of W_n via linear regression of observed order statistics on the means of the standard normal order statistics, and the fact that the linear regression line is very sensitive to outliers or observations corresponding to extreme values in the x-direction, we can expect that the SW test is sensitive (or powerful) for detecting the non-normality due to outliers, or against a density with heavier tail than the normal density (e.g., the Cauchy distribution). These features of the SW test are quite easy to understand even for most biomedical researchers.

2.2 The skewness and kurtosis tests

The skewness and kurtosis have long been suggested for detecting non-normality in the univariate setting [17]. For general multivariate data, Mardia [16] constructed two statistics for measuring multivariate skewness and kurtosis. Using the notation for standardized data in (1), the skewness statistic MS is:

$$\mathrm{MS} = \frac{1}{6n} \sum_{i,j=1}^{n} \left(Y_i^{\mathrm{T}} Y_j \right)^3. \quad (3)$$

The kurtosis statistic MK is:

$$\mathrm{MK} = \sqrt{\frac{n}{8p(p+2)}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \|Y_i\|^4 - \frac{p(p+2)(n-1)}{n+1} \right\}. \quad (4)$$

The skewness test rejects the hypothesis of normality if MS is too large, and the test based on the centralized kurtosis statistic MK rejects the null hypothesis of normality if its absolute value |MK| is too large, that is, it exceeds the appropriate critical value. Both the skewness and kurtosis tests are simple and informative [4], and provide specific information about non-normality of the data. In the univariate case, D'Agostino and Pearson [5] proposed the K^2 test, by combining the skewness and kurtosis measures. Bowman and Shenton [2] considered an alternative way of combining skewness and kurtosis equivalent to the following statistic

$$\mathrm{MSK} = \mathrm{MS} + |\mathrm{MK}|^2, \quad (5)$$

which has an asymptotic χ_2^2 distribution in the univariate case. In the multivariate case, Doornik and Hansen [7] demonstrated some practical utility and good power performance of MSK, especially against generalized Burr-Pareto-logistic distribution with normal marginals [7, Table 4]. However, these tests are not consistent for testing general alternatives and can have very low power against many alternatives.

2.3 The Henze-Zirkler test

Extending the work of [1, 10], Henze and Zirkler [13] proposed a test for normality using

$$\mathrm{HZ}_\beta = n(4I_E + D_{n,\beta} I_{E^c}), \quad (6)$$

where $\beta \in \mathrm{R}$, I_E, I_{E^c} are indicator functions with $E = \{S_n \text{ is singular}\}$ and in terms of Y_i in (1),

$$D_{n,\beta} = n^{-2} \sum_{j,k=1}^{n} \exp(-\beta^2 \|Y_j - Y_k\|^2/2) + \left(1 + 2\beta^2\right)^{-p/2} - 2\left(1 + \beta^2\right)^{-p/2} n^{-1} \sum_{j=1}^{n} \exp\left[-\beta^2 \|Y_j\|^2 / \left\{ 2(1+\beta^2) \right\} \right].$$

The Henze-Zirkler (HZ) test rejects normality if HZ_β is too large. They also proposed an optimal choice of the parameter β in using HZ_β in the p-variate case as

$$\beta^* = 2^{-1/2} \{ [(2p+1)n]/4 \}^{1/(p+4)}.$$

A drawback of the HZ test is that, when H_0 is rejected, the possible violation of normality is generally not clear. Thus many biomedical researchers would prefer an more informative and equally or more powerful test than the HZ test.

3. The new test

The main goal here is to construct a simple test with a combination of easy-to-understand components. Starting from univariate data, there have been many known tests obtained via combinations of kurtosis with other tests, among which the most well-known one [5] combines the kurtosis statistic MK with the skewness statistic MS. However, MS is powerful only when the alternative is skewed, thus it generally lacks power to test against non-skewed alternatives. On the other hand, SW has excellent power to detect skewed alternatives. Moreover, SW is also consistent against general alternatives [14]. Thus, we propose to combine kurtosis with SW, instead of with MS.

For a p-dimensional random vector Y, it is well known that Y is normal if and only if $\theta^T Y$ is univariate normal for all $\theta \in \{\theta \in R^p: \|\theta\| = 1\}$. It can also be proved that, if Y is not normal, the event that $\theta^T Y$ is univariate normal can only occur for θ in a zero measure subset of $\{\theta \in R^p: \|\theta\| = 1\}$, which is a null set if one picks a θ randomly as discussed in [19, Theorem 1]. Thus it seems reasonable to pick a few data-driven directions, say $\theta \in \Theta$, and use SW test to detect non-normality in univariate projections on those directions. More specifically, for multivariate data $X_1,\ldots,X_n$, using the standardization discussed in (1) to obtain $\{Y_i\}_{1 \leq i \leq n}$, we can then consider the the following statistic based on projection of Y_i's in the direction θ:

$$G_n(\theta) = W_n(\theta^T Y_1, \ldots, \theta^T Y_n), \quad (7)$$

where W_n is the Shapiro-Wilk function in (2). Fattorini [11] considered a test which rejects normality for small values of following statistic

$$FA = \min_{1 \leq i \leq n} G_n(Y_i). \quad (8)$$

The Fattorini (FA) test is recommended by [23, pp.220] for testing multivariate normality in addition to the HZ test. It is clear that the FA test is based on detecting non-normality of multivariate data in the most "extreme" directions corresponding to the smallest G_n values evaluated at random directions $\{\|Y_j\|^{-1}Y_j\}_{1 \leq j \leq n}$.

More generally, it seems natural to detect non-normality of the p most "extreme" directions corresponding to p smallest G_n values evaluated at random directions $\{\|Y_j\|^{-1}Y_j\}_{1 \leq j \leq n}$, denoted Θ_1. It is also reasonable to consider the p marginal variates which often have specific physical meanings and are of interest to practitioners. As a result, our new test statistic for normality is:

$$T_n = 1 - (2p)^{-1} \sum_{\theta \in \Theta_1 \cup \Theta_2} G_n(\theta) I_A, \quad (9)$$

where G_n is the function in (7), Θ_1 consists of the p most "extreme" directions corresponding to p smallest G_n values evaluated at random directions $\{\|Y_j\|^{-1}Y_j\}_{1 \leq j \leq n}$ and $\Theta_2 = \{e_j\}_{1 \leq j \leq p}$ with $e_j = (0,\ldots,0, 1, 0,\ldots,0)^T$ (the unit vector with its j-th opponent being 1 and all others being 0), and

$$A = \{c_1 \leq MK \leq c_2\}, \quad (10)$$

with c_1, c_2 being certain percentiles of MK in (4) under H_0, e.g., c_1 and c_2 are 1% and 99% quantiles of MK as used in Section 4 for power simulation.

Note that, large values of $1 - W_n$, |MK| all indicate departure from normality, thus the new test will reject H_0 when T_n in (9) is too large. The statistic T_n is easy to evaluate using functions in R for W_n. Simulation results reported in the next section show that the new test has competitive power in all dimensions compared to some best known tests.

4. Numerical Studies

For power comparison of our proposed test T_n in (9), competing tests considered here include the Shapiro-Wilk test W_n in (2) (univariate case), the MSK test in (5), the Fatorrini test FA in (8), and the Henze-Zirkler test HZ_β in (6) with the optimal choice of β in [13]. A broad range of alternative models are considered in the simulation. In the univariate case, the alternatives include those from Table 3 of the classic goodness-of-fit paper by Stephens [22], and also the Pearson type II, Pearson type VII distributions, and some other spherically symmetric distributions as considered in [13]. In the multivariate cases, alternative distributions include those considered in Table 6.4 in [13] and other similar ones. Significance levels of all tests are set at α = 5%.

We mention briefly in Section 2 that the null critical values can be simulated using empirical distribution via the Monte Carlo approach. This is due to the following two facts: First of all, under H_0, as functions of the standardized data in (1), all the test statistics considered here have null distributions that do not depend on the location vector and the covariance matrix [13]. Thus, we can simulate their null distributions directly using the zero location vector and identity covariance matrix; Secondly, the empirical distribution function $F_m(t)$ for the Monte Carlo sample (of size m) is uniformly (in the Kolmogorov-Smirnov distance) close to the true cumulative distribution function $F(t)$. In fact, by the well-known Dvoretzky-Kiefer-Wolfowitz inequality [18, Lemma 5.1],

$$Pr(\sup_{t\in R} |F_m(t) - F(t)| > \sqrt{m^{-1}\log m}) \le Cm^{-2},$$

where C is a constant independent of F. Then, by the Borel-Cantelli lemma, almost surely, we have $\sup_t |F_m(t) - F(t)| \le \sqrt{m^{-1}\log m}$. Note that m is the number of replicates in Monte Carlo, which can easily be chosen as 5 million or more if needed. Therefore, the Monte Carlo approximated critical values, calculated as quantiles of F_m, can be as accurate as desired by choosing a sufficiently large m.

For $HZ_{n,\beta}$, Henze and Zirkler [13] developed an optimal choice β^* for β, where $\beta^* = 2^{-1/2}\{[(2p + 1)n]/4\}^{1/(p+4)}$. Different β values in (6) yield tests that are sensitive to different types of alternatives. Simulations indicated that β^* results in a test having good power against a broad range of alternatives, thus as suggested in [13], it is a preferred choice of β as an omnibus test for normality here. We use the 1% and 99% empirical quantiles of MK under H_0, obtained through Monte Carlo simulation, for c_1, c_2 respectively. For example, when $n = 50$, $p = 2$, $c_1 = -1.455$, $c_2 = 2.551$ Note that in this way, for the level α = 5% test, the critical value of T_n is less than 1. Critical values of all test statistics are calculated using 100,000 simple random samples from the p-variate standard normal distribution. Those critical values are then used to calculate the empirical powers in Table A1-A4, and are given in Table 1. In addition, for practical convenience, the critical values of all test statistics considered corresponding to the level α = 5% test are presented in Table A5 for more sample size choices.

Empirical powers are obtained based on the percentage of 5,000 Monte Carlo samples declared significant. Let θ be the true power, the standard error of the empirical power is then $\sqrt{\theta(1-\theta)/5000} \leq \sqrt{1/20000} = 0.007$. Power comparison results corresponding to sample size $n = 50$ are reported. In addition, average ranks are appended at the end of each power table, where smaller values indicate better power performance. Average power across the alternatives are also included for comparison.

For all the critical values used in the empirical power calculation, the corresponding empirical type I errors, calculated as the proportion of rejection using 100,000 Monte Carlo normal samples, are also provided in Table 1. In the meanwhile, those type I errors are visualized in Figure 1. Power results of the univariate case are summarized in Table A1, where NormMix(p, μ_2, σ_2) stands for the univariate normal mixture distribution whose density is given by $p\varphi(x) + (1 - p)\varphi((x - \mu_2)/\sigma_2)$, with $\varphi(x)$ being the density of standard normal distribution. In this case, the FA test reduces to W_n since the Shapiro-Wilk statistic is location-scale invariant. From Table A1, even in the univariate case, the new test T_n is never seriously less powerful than W_n for all alternatives investigated. Moreover, T_n can be significantly more powerful than W_n for several alternatives considered here. For examples, T_n is 60% more powerful than W_n for Beta(2, 2) and PSII(1). Whether measuring by average rank or average power, T_n slightly outperforms the benchmark Shapiro-Wilk W_n test in the univariate case. Also, as demonstrated in Table A1, T_n is generally more powerful than the HZ test and MSK test.

In the multivariate case, $F_1 \otimes F_2$ denotes the distribution with independent marginal distributions F_1 and F_2, and the product of k independent copies of F_1 will be denoted by F_1^k. Table A2 displays the power performances in the bivariate case. MVNMIX(a. b. c. d) stands for the multivariate normal mixture distribution with density $aN(0,\Sigma_1) + (1 - a)N(b\mathbf{1},\Sigma_2)$, where $\mathbf{1}$ is the column vector with all elements being 1, $\Sigma_1 = (1 - c)I + c\mathbf{1}\mathbf{1}^T$ and $\Sigma_2 = (1 - d)I + d\mathbf{1}\mathbf{1}^T$. Table A3 and Table A4 show results of testing 5 and 10 dimensional normality. It is clear that the new test outperforms the MSK, FA, HZ tests for an overwhelming majority of the examined alternatives. The new test also has higher power than the HZ test in more than 90% of alternatives studied, and the power difference are often quite substantial. Overall, the new test T_n is superior to HZ in terms of both average power and average rank. For example, when $p = 5$, the new test T_n is the most powerful among the four competitors for testing 46 out of the 48 alternatives, and in the two alternatives the new test is second most powerful. Thus T_n by far has the best average rank. When the dimension $p = 10$, the average power of T_n among these 48 alternatives is 73.04% which is far above the 47.9% average power of the HZ test. As illustrated using these alternatives, the power advantage of T_n over the best known competitor HZ test is overwhelming. The overall power advantage of the new test over MSK and FA is even more striking.

5. Examples

5.1 The Ramus Bone Data

Elston and Grizzle [9] gave a very interesting data of ramus height (in millimeters) of 20 boys each measured at age 8, 8.5, 9 and 9.5 years old. The data is fairly well known because it has a few interesting features. The recorded heights of the ramus bone marginally appear to be normally distributed while not jointly normal [24, pp. 126-130]. Specifically, we consider testing for bivariate normality of the ramus height data corresponding to the measurements at ages 8 and 9. The p-values corresponding to MSK, FA, HZ, T_n are 0.146, 0.017, 0.002, 0.037 respectively. Thus, at level $\alpha = 5\%$, our test, FA test and HZ test reject the null hypothesis of bivariate normality, while MSK fails to reject H_0. On the other hand, if we test the multivariate normality of ages 8, 8.5, 9, 9.5, the p-values, for MSK, FA, HZ,

T_n are 0.002, 0.054, < 0.001, < 0.001 respectively. The potential reasons for the non-normality include the latent genetic heterogeneities which lead to different growth profiles. For example, Timm [24] pointed out that observation 9 appears to be an outlier, that is, the #9 boy's growth profile appears to differ from many others.

5.2 Fisher's Iris Data

Fisher's Iris data set is a multivariate data set introduced by Fisher [12] to demonstrate the use of unclassified observations in estimating discriminant function. Looney [15] considered testing univariate normality of each of the four measurements (sepal length, sepal width, petal length, and petal width) corresponding to variety *Iris setosa*, and found that normality assumption is tenable for all the variables except the petal width. Consider the multivariate normality of the four measurements, When MSK, FA, HZ, T_n are applied to all the 150 observations, consisting of all three species: *Iris setosa, Iris versicolor, Iris virginica*, the multivariate normality of the four measurements is rejected by all tests (all p-values are less than 2%). If we only use the first 50 observations, corresponding to *Iris setosa*, to test the four dimensional normality, we obtain corresponding *p*-values of 0.085, 0.065, 0.049, 0.037 for MSK, FA,HZ, T_n respectively. Thus at level $\alpha = 5\%$, the HZ test and our test T_n are able to reject the normality assumption of the four measurements on variety *Iris setosa* even with only 1/3 of the data. Rejecting multivariate normality for this data is consistent with the findings in the literature [21].

6. Conclusion

This paper presents a new test for normality that is very competitive in terms of power in both univariate and multivariate cases. As demonstrated by the numerical study, the new test outperforms the best known competitors in the literature, and the power advantage of the new test over its best known competitors is overwhelming in high dimensions. In addition, the practical utility of the proposed test is also illustrated using two well-known real data examples. Moreover, our projection-based statistic is very easy to understand and to implement using existing freely available numerical tools such as the `shapiro.test` function in the R computing platform. The R program for computing the p-value of the proposed test will be freely available at authors's websites upon publication of the manuscript.

Appendix A. Power Tables

Table A1

Empirical power of tests for univariate normality at significance level $\alpha = 5\%$ when $n = 50$.

Alternatives	MSK	W_n	HZ	T_n
N(0,1)	5.0	4.8	5.1	4.7
Exp(1)	96	100	99	100
Lognormal(0,1)	100	100	100	100
Lognormal(0,0.5^2)	83	93	84	91
Gamma(5,1)	46	59	45	55
$\chi^2(1)$	100	100	100	100
$\chi^2(2)$	97	100	99	100
$\chi^2(5)$	72	89	78	88
$\chi^2(10)$	47	59	46	57
Cauchy(0,1)	100	100	100	100

Alternatives	MSK	W_n	HZ	T_n
$t(2)$	88	86	84	86
$t(5)$	44	35	26	36
Beta(1,1)	0	76	62	84
Beta(1,2)	14	83	70	80
Beta(2,2)	0	15	17	25
Logistic(0,1)	26	19	14	19
Halfnormal	59	94	80	92
Weibull(0.8)	100	100	100	100
Weibull(1)	97	100	99	100
Weibull(1.5)	63	88	73	86
NormMix(.5,4,1)	0	90	96	94
NormMix(.5,3,1)	0	38	51	49
NormMix(.5,5,1)	2	100	100	100
NormMix(.5,6,3)	35	99	98	98
NormMix(.9,6,3)	98	98	96	98
NormMix(.5,2,1/3)	36	99	98	98
NormMix(.9,2,1/3)	1	8	10	9
PSII(0)	0	76	62	85
PSII(1)	0	14	16	24
PSVII(4)	31	23	15	23
PSVII(5)	24	17	11	17
SPH(Gamma(5,1))	0	94	98	94
SPH(Beta(1,1))	0	75	62	84
SPH(Beta(1,2))	0	5	6	7
SPH(Beta(2,2))	1	99	99	100
Average Power	42.92	71.47	67.5	72.85
Average Rank	3.059	1.676	2.529	1.529

Table A2

Empirical power of tests for bivariate normality at significance level $\alpha = 5\%$ when $n = 50$.

Alternatives	MSK	FA	HZ	T_n
$N(0,1)^2$	5.2	5.3	5.2	4.8
$Exp(1)^2$	100	100	100	100
$Lognormal(0,1)^2$	100	100	100	100
$Lognormal(0, .5^2)^2$	96	97	94	99
$Gamma(.5,1)^2$	100	100	100	100
$Gamma(5,1)^2$	60	64	52	71
$\chi^2(1)^2$	100	100	100	100
$\chi^2(2)^2$	100	100	100	100
$\chi^2(5)^2$	88	93	86	96

Alternatives	MSK	FA	HZ	T_n
$\chi^2(10)^2$	60	63	51	70
$\chi^2(15)^2$	44	45	35	51
Cauchy$(0,1)^2$	100	100	100	100
$t(2)^2$	96	95	95	97
$t(5)^2$	55	46	33	52
Logistic$(0,1)^2$	30	23	14	26
Beta$(1,1)^2$	0	53	67	90
Beta$(1,2)^2$	17	76	80	91
Beta$(2,2)^2$	0	6	19	33
Halfnormal^2	76	93	87	97
Weibull$(.8)^2$	100	100	100	100
Weibull$(1)^2$	100	100	100	100
Weibull$(1.5)^2$	79	89	83	95
$N(0,1)\otimes$Exp(1)	89	99	93	99
$N(0,1)\otimes\chi^2(5)$	58	75	52	77
$N(0,1)\otimes t(5)$	31	27	17	29
$N(0,1)\otimes$Beta$(1,1)$	1	33	31	43
$N(0,1)\otimes$Beta$(1,2)$	7	51	41	54
MVNMIX$(.5,2,0,0)$	2	10	20	12
MVNMIX$(.5,4,0,0)$	1	100	100	98
MVNMIX$(.5,2,.9,0)$	73	67	81	62
MVNMIX$(.5,.5,.9,0)$	34	39	34	28
MVNMIX$(.5,.5,.9,-.9)$	72	90	93	74
MVNMIX$(.7,2,.9,.3)$	76	73	66	75
MVNMIX$(.3,1,.9,-.9)$	97	100	99	99
PSII(0)	0	27	71	96
PSII(1)	0	6	24	52
PSVII(2)	98	97	97	98
PSVII(3)	75	67	59	71
PSVII(5)	41	32	19	34
SPH$($Exp$(1))$	98	98	100	99
SPH$($Gamma$(5,1))$	4	8	18	13
SPH$($Beta$(1,1))$	0	3	15	4
SPH$($Beta$(1,2))$	24	29	66	31
SPH$($Beta$(2,2))$	0	3	10	24
Average Power	55.35	64.59	65.12	70.63
Average Rank	2.651	2.186	2.349	1.512

Table A3

Empirical power of tests for 5 dimensional normality at level $\alpha = 5\%$ when $n = 50$.

Alternatives	MSK	FA	HZ	T_n
$N(0,1)^5$	5.0	5.1	5.1	5.0
$Exp(1)^5$	100	100	100	100
Lognormal$(0,1)^5$	100	100	100	100
Lognormal$(0,.5^2)^5$	99	96	96	100
Gamma$(.5,1)^5$	100	100	100	100
Gamma$(5,1)^5$	69	50	48	86
$\chi^2(1)^5$	100	100	100	100
$\chi^2(2)^5$	100	100	100	100
$\chi^2(5)^5$	96	82	88	100
$\chi^2(10)^5$	69	50	48	88
$\chi^2(15)^5$	48	35	28	66
Cauchy$(0,1)^5$	100	100	100	100
$t(2)^5$	100	99	99	100
$t(5)^5$	68	59	32	74
Logistic$(0,1)^5$	38	29	14	40
Beta$(1,1)^5$	0	0	52	86
Beta$(1,2)^5$	6	7	68	87
Beta$(2,2)^5$	0	0	14	35
Halfnormal^5	82	58	84	100
Weibull$(.8)^5$	100	100	100	100
Weibull$(1)^5$	100	100	100	100
Weibull$(1.5)^5$	88	67	80	99
$N(0,1)\otimes Exp(1)^4$	100	99	100	100
$N(0,1)\otimes\chi^2(5)^4$	88	74	74	98
$N(0,1)\otimes t(5)^4$	57	51	24	64
$N(0,1)\otimes Beta(1,1)^4$	0	1	39	58
$N(0,1)\otimes Beta(1,2)^4$	5	6	54	67
$N(0,1)^3\otimes Exp(1)^2$	90	89	81	99
$N(0,1)^3\otimes\chi^2(5)^2$	53	50	35	76
$N(0,1)^3\otimes t(5)^2$	33	32	13	37
$N(0,1)^3\otimes Beta(1,1)^2$	1	3	17	16
$N(0,1)^3\otimes Beta(1,2)^2$	5	6	25	23
MVNMIX$(.5,2,0,0)$	3	4	22	8
MVNMIX$(.5,4,0,0)$	2	10	70	11
MVNMIX$(.5,2,.9,0)$	100	89	100	100
MVNMIX$(.5,.5,.9,0)$	92	76	99	94
MVNMIX$(.5,.5,.9,-.1)$	94	83	99	96
MVNMIX$(.7,2,.9,.3)$	100	96	99	100
MVNMIX$(.3,1,.9,-.1)$	86	76	91	89

Alternatives	MSK	FA	HZ	T_n
PSII(0)	0	0	65	100
PSII(1)	0	0	30	88
PSVII(4)	99	97	98	99
PSVII(5)	92	82	71	92
PSVII(10)	38	29	13	33
SPH(Exp(1))	100	100	100	100
SPH(Gamma(5,1))	70	47	44	68
SPH(Beta(1,1))	83	45	100	92
SPH(Beta(1,2))	100	95	100	100
SPH(Beta(2,2))	37	13	71	42
Average Power	64.38	57.97	68.41	79.4
Average Rank	2.125	2.896	2.208	1.229

Table A4

Empirical power of tests for 10 dimensional normality at level α = 5% when n = 50.

Alternatives	MSK	FA	HZ	T_n
Normal^{10}	5.1	4.8	4.9	5.1
$\mathrm{Exp}(1)^{10}$	100	96	100	100
$\mathrm{Cauchy}(0,1)^{10}$	100	100	100	100
$\mathrm{Gamma}(.5,1)^{10}$	100	100	100	100
$\mathrm{Gamma}(5,1)^{10}$	64	36	27	94
$\mathrm{Lognormal}(0,.5^2)^{10}$	100	88	87	100
$\mathrm{Lognormal}(0,1)^{10}$	100	100	100	100
$\chi^2(2)^{10}$	100	97	100	100
$\chi^2(5)^{10}$	94	63	65	100
$\chi^2(10)^{10}$	62	35	26	94
$\chi^2(15)^{10}$	43	25	16	76
$t(2)^{10}$	100	100	99	100
$t(5)^{10}$	77	62	20	89
$\mathrm{Logistic}(0,1)^{10}$	39	25	8	50
$\mathrm{Beta}(1,1)^{10}$	0	0	27	72
$\mathrm{Beta}(1,2)^{10}$	2	2	35	83
$\mathrm{Beta}(2,2)^{10}$	0	1	11	26
$\mathrm{Halfnormal}^{10}$	69	29	52	100
$\mathrm{Weibull}(.8)^{10}$	100	100	100	100
$\mathrm{Weibull}(1)^{10}$	100	96	100	100
$\mathrm{Weibull}(1.5)^{10}$	82	42	51	100
$N(0,1)^9 \otimes \mathrm{Exp}(1)$	27	29	11	49
$N(0,1)^9 \otimes \chi^2(5)$	13	13	7	23
$N(0,1)^9 \otimes t(5)$	12	14	6	15

Alternatives	MSK	FA	HZ	T_n
$N(0,1)^9 \otimes Beta(1,1)$	3	4	6	6
$N(0,1)^9 \otimes Beta(1,2)$	5	5	7	6
$N(0,1)^9 \otimes Beta(2,2)$	4	4	5	5
$N(0,1)^5 \otimes Exp(1)^5$	96	79	70	100
$N(0,1)^5 \otimes \chi^2(5)^5$	60	40	24	95
$N(0,1)^5 \otimes t(5)^5$	43	39	9	55
$N(0,1)^5 \otimes Beta(1,1)^5$	0	2	12	16
$N(0,1)^5 \otimes Beta(1,2)^5$	3	4	15	22
$N(0,1)^5 \otimes Beta(2,2)^5$	1	3	7	8
$N(0,1)^2 \otimes Exp(1)^8$	100	92	96	100
$N(0,1)^2 \otimes \chi^2(5)^8$	87	56	46	100
$N(0,1)^2 \otimes t(5)^8$	65	53	15	79
$N(0,1)^2 \otimes Beta(1,1)^8$	0	1	20	45
$N(0,1)^2 \otimes Beta(1,2)^8$	2	2	26	57
$N(0,1)^2 \otimes Beta(2,2)^8$	0	1	9	15
PSII(0)	0	0	48	100
PSII(1)	0	0	29	95
PSII(4)	0	0	11	48
PSVII(6)	100	100	100	100
PSVII(8)	98	89	71	98
PSVII(10)	86	62	29	84
SPH(Gamma(5,1))	100	88	96	100
SPH(Beta(1,1))	100	86	100	100
SPH(Beta(1,2))	100	100	100	100
SPH(Beta(2,2))	99	60	100	100
Average Power	54.91	46.28	47.9	73.04
Average Rank	2.188	2.854	2.521	1.042

Table A5

Critical values for the level $\alpha = 5\%$ test of p dimensional normality

Sample Size	$p = 1$	$p = 2$	$p = 5$	$p = 10$
		MSK		
n = 20	4.291	9.190	42.886	211.679
n = 25	4.682	9.661	45.106	222.915
$n = 30$	4.856	10.055	46.679	230.264
$n = 35$	5.017	10.322	47.661	235.287
$n = 40$	5.187	10.487	48.504	239.648
$n = 45$	5.310	10.683	48.955	242.658
$n = 50$	5.329	10.712	49.509	244.326

Sample Size	$p = 1$	$p = 2$	$p = 5$	$p = 10$
		FA		
$n = 20$	0.096	0.141	0.244	0.463
$n = 25$	0.080	0.119	0.199	0.372
$n = 30$	0.070	0.103	0.168	0.309
$n = 35$	0.061	0.091	0.146	0.266
$n = 40$	0.055	0.081	0.129	0.234
$n = 45$	0.05	0.074	0.116	0.208
$n = 50$	0.046	0.068	0.106	0.187
		HZ		
$n = 20$	0.546	0.728	0.910	0.985
$n = 25$	0.591	0.766	0.924	0.987
$n = 30$	0.624	0.798	0.936	0.989
$n = 35$	0.658	0.820	0.944	0.991
$n = 40$	0.685	0.840	0.952	0.992
$n = 45$	0.717	0.861	0.957	0.992
$n = 50$	0.729	0.873	0.963	0.993
		T_n		
$n = 20$	0.1001	0.1092	0.1018	0.1490
$n = 25$	0.0842	0.0927	0.0864	0.1185
$n = 30$	0.0729	0.0811	0.0759	0.0998
$n = 35$	0.0644	0.0719	0.0681	0.0865
$n = 40$	0.0581	0.0644	0.0618	0.0769
$n = 45$	0.0526	0.0589	0.0567	0.0694
$n = 50$	0.0483	0.0539	0.0525	0.0632

References

1. Baringhaus L, Henze N. A consistent test for multivariate normality based on the empirical characteristic function. Metrika. 1988; 35:339–348.

2. Bowman KO, Shenton LR. Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2. Biometrika. 1975; 62:243–250.

3. Coin D. A goodness-of-fit test for normality based on polynomial regression. Comput. Statist. Data Anal. 2008; 52:2185–2198.

4. D'Agostino RB, Belanger A, D'agostino RB Jr. A suggestion for using powerful and informative tests of normality. The American Statistician. 1990; 44:316–321.

5. D'Agostino RB, Pearson ES. Tests for departure from normality. emprical results for the distribution of b_2 and $\sqrt{b_1}$. Biometrika. 1973; 62:613–622.

6. D'Agostino, RB.; Stephens, MA. Goodness-of-Fit Techniques. New York: Marcel Dekker, Inc.; 1986.

7. Doornik JA, Hansen H. An omnibus test for univariate and multivariate normality. Oxford Bulletin of Economics and Statistics. 2008; 70:927–939.

8. Eaton ML, Perlman MD. The non-singularity of generalized sample covariance matrices. Ann. Statist. 1973; 1:710–717.

9. Elston RC, Grizzle JE. Estimation of time-response curves and their confidence bands. Biometrics. 1962; 18:148–159.

10. Epps TW, Pulley LB. A test for normality based on the empirical characteristic function. Biometrika. 1983; 70:723–726.

11. Fattorini L. Remarks on the use of the shapiro-wilk statistic for testing multivariate normality. Statistica. 1986; 46:209–217.

12. Fisher RA. The use of multiple measurements in taxonomic problems. Annals of Eugenics. 1936; 7:179–188.

13. Henze N, Zirkler B. A class of invariant consistent tests for multivariate normality. Comm. Statist. Theory Methods. 1990; 19:3595–3617.

14. Leslie J, Stephens MA, Fotopoulos S. Asymptotic distribution of the shapiro-wilk w for testing for normality. Ann. Statist. 1986; 14:1497–1506.

15. Looney SW. How to use tests for univariate normality to assess multivariate normality. The American Statistician. 1995; 49:64–70.

16. Mardia KV. Measures of multivariate skewness and kurtosis with applications. Biometrika. 1970; 57:519–530.

17. Pearson ES. A further development of tests for normality. Biometrika. 1930; 22:239–249.

18. Shao, J. Mathematical Statistics. 2nd ed.. Springer-Verlag New York Inc; 2003.

19. Shao Y, Zhou M. A characterization of multivariate normality through univariate projections. Journal of Multivariate Analysis. 2010; 101:2637–2640.

20. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika. 1965; 52:591–611.

21. Small N. Marginal skewness and kurtosis in testing multivariate normality. Applied Statistics. 1980; 29:85–87.

22. Stephens MA. EDF statistics for goodness of fit and some comparisons. Journal of the American Statistical Association. 1974; 69:730–737.

23. Thode, HC, Jr. Testing for Normality. New York: Marcel Dekker, Inc.; 2002.

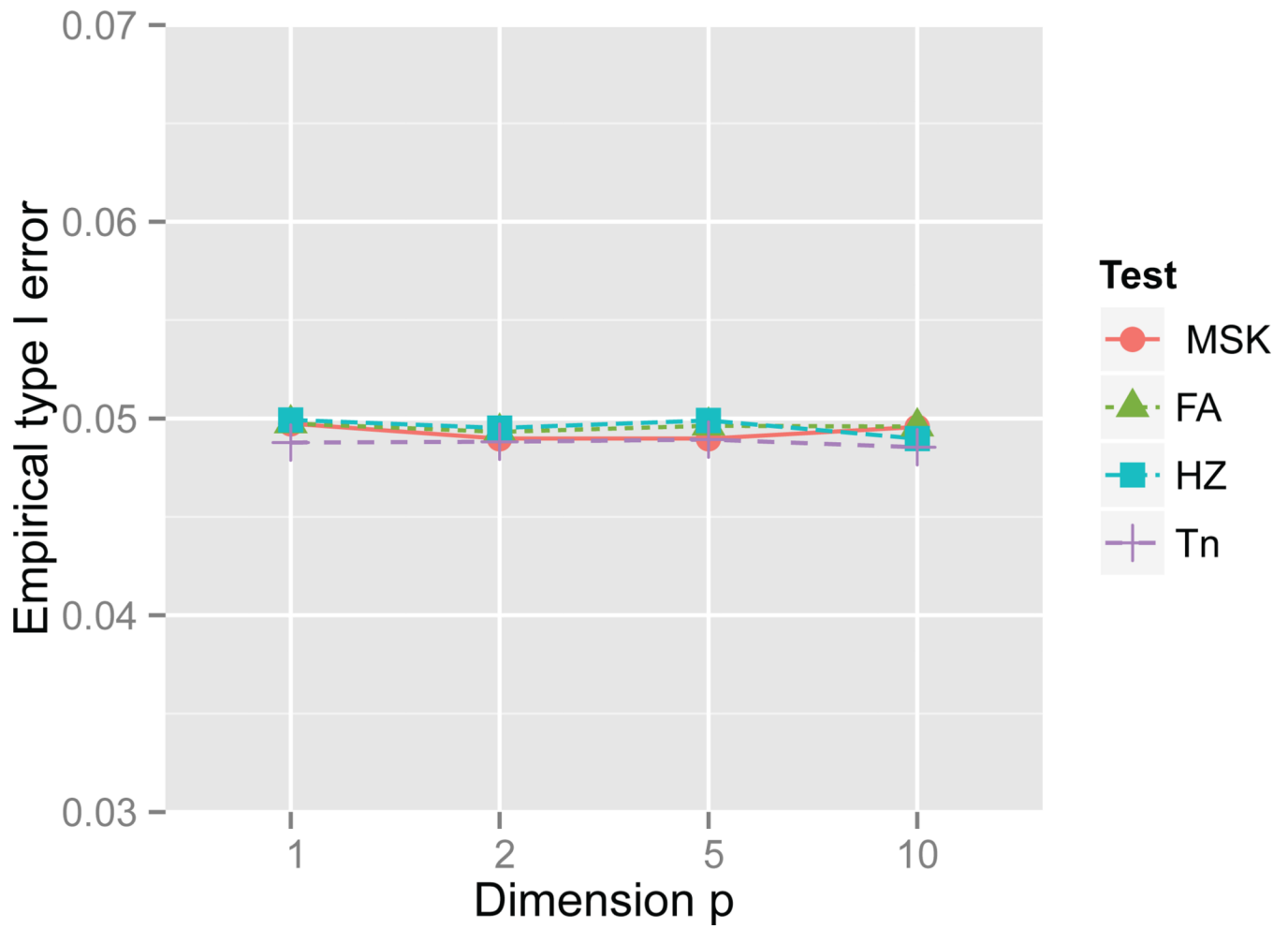24. Timm, NH. Applied multivariate analysis. New York: Springer; 2002.

Figure 1.
Empirical type I error of tests for normality at significance level α = 5% when *n* = 50.

Table 1

Empirical critical values and corresponding type I errors for the level $\alpha = 5\%$ test of p dimensional normality when $n = 50$.

Dimension	Critical Value				Type I Error			
	MSK	FA	HZ	T_n	MSK	FA	HZ	T_n
$p = 1$	5.329	0.046	0.729	0.048	0.0497	0.0498	0.0499	0.0488
$p = 2$	10.712	0.068	0.873	0.054	0.0490	0.0493	0.0495	0.0488
$p = 5$	49.509	0.106	0.963	0.053	0.0490	0.0496	0.0499	0.0489
$p = 10$	244.326	0.187	0.993	0.063	0.0496	0.0496	0.0490	0.0485