British Journal of Applied Science & Technology
16(1): 1-8, 2016, Article no.BJAST.26256
ISSN: 2231-0843, NLM ID: 101664541

SCIENCEDOMAIN ***international***
www.sciencedomain.org

Usage of Multidimensional Scaling Technique for Evaluating Performances of Multivariate Normality Tests

Soner Yiğit[1] and Mehmet Mendeş[1*]

[1]Department of Animal Science, Faculty of Agriculture, Biometry and Genetics Unit, Çanakkale Onsekiz Mart University, 17100, Çanakkale, Turkey.

Authors' contributions

This work was carried out in collaboration between both authors. Author MM designed the simulation study, wrote the simulation codes and wrote the first draft of the manuscript. Author SY wrote some part of simulation codes and managed literature searches. Both authors read and approved the final version of the manuscript.

Article Information

DOI: 10.9734/BJAST/2016/26256
Editor(s):
(1) Vitaly Kober, Department of Computer Science, CICESE, Mexico.
Reviewers:
(1) Sherin Zafar, Jamia Hamdard University, India.
(2) Ahmet Sayar, Kocaeli University, Turkey.
(3) R. Praveen Sam, Jawaharlal Nehru Technological University, Anantapur, India.
Complete Peer review History: http://sciencedomain.org/review-history/14595

Original Research Article

Received 7th April 2016
Accepted 3rd May 2016
Published 12th May 2016

ABSTRACT

This simulation study has been conducted to evaluate the performances of six different multivariate normality tests under different experimental conditions. Obtained results of 50,000 Monte Carlo Simulation showed the most reliable when the Royston (Roy), Srivastava-Hui (S-H), and Doornik-Hansen test (D-H) have been applied. The above mentioned tests retained Type I error rates at nominal alpha level (0.05). Whereas, the estimations of Type I error of Mardia's Skewness (M-S), Mardia's Kent (M-K) and Henze and Zirkler (H-Z) test caused variations depending on sample size and number of variables. The estimations of test power of all tests have been affected by distribution shape, and the all related tests produced highly test power values especially when samples were taken from Multivariate Cauchy and Lognormal distributions. On the other hand, the estimations of test power of all tests have been found extremely low when samples were taken from multivariate t-distribution with 10 d.f. Multidimensional Scaling (MDS) technique has been applied to classify the tests those have had similar performance and the factors those affected the

**Corresponding author: E-mail: mmendes@comu.edu.tr;*

performances of the above mentioned tests. At the end of Multidimensional Scaling analyses, it has been observed that the Roy, S-H and D-H tests showed similar performance, and the performances of these tests were obviously different than that of the others in general.

Keywords: Multivariate normality; type I error; test power; multidimensional scaling technique.

1. INTRODUCTION

Practically, studying with multivariate data sets is very common in case of those studies which related to the fields of medicine, agriculture, forestry, aquaculture, sociology, psychology, and education. Multivariate statistical techniques such as MANOVA, Discriminant Analysis, Multivariate Regression are commonly used in practice for the assumption of multivariate normality (MVN). The sensitivity of these multivariate techniques to the MVN has been reported by different researchers, and many multivariate normality tests have been suggested to check out this assumption whether met or not. This assumption has generally been ignored [1-3]. Consequently, the researchers may be come up with conflicting or unreliable results. Many simulation studies have been carried out by the following researchers and they emphasized on the importance of the MVN assumption for these methods and also described the effect of deviations from multivariate normality on reliability of results [1,2,4-13].

Many different multivariate normality tests have been proposed and there are several simulation studies in the literature cited for comparing the performances of some of those tests [1,2,5,7,8,10-14]. When these tests are compared in terms of their performances then it would be more informative if both of Type I error and test power estimates of these tests are considered simultaneously. Therefore, it is an important issue to classify these tests on the basis of their performance by using a graphical methods namely Multidimensional Scaling Technique (MDS). After that it would be possible to see the tests that have similar or same performances under different experimental conditions. For this purpose, we think that it might be a good choice to classify the all MVN tests considered based on their performance as a subjective hypothesis testing because there is no any other hypothesis test to compare the results of simulation studies (Type I error rate and test power).

2. MATERIALS AND METHODS

The main objective of this study is compare the type I error rate and test power of different multivariate normality tests (MVN) and also classifying these tests on the basis of their performances. First of all, a Monte Carlo simulation study has been designed for this purpose. In this study, a total of 50,000 data sets along with the combinations of sample size i.e.; n=20, 30, 50 and 100; and number of variables i.e.; p=2, 3, 4 and 5 have been generated from four different multivariate distributions ranging from the multivariate normal to severe deviations from multivariate normality. True correlation between the variables has been determined as Rho=0.60. A total of 6 different tests of the MVN namely: Mardia's Skewness (M-S), Royston (Roy), Srivastava-Hui (S-H), Doornik-Hansen Test (D-H), Henze-Zirkler Test (H-Z), and Mardia's Kent (M-K) have been kept in consideration while carrying out this study. Since it is possible to find more detail information from both different multivariate text books and articles in this regard, but we did not feel it necessary whether to provide theoretical information related to these tests. Moreover, the above mentioned tests have also been classified based on their performance by using Multidimensional Scaling Technique (MDS).

3. RESULTS

3.1 Results of Type I Error Rates

The empirical type I error rates regarding to the six tests are given in Table 1. Table 1 showed that the type I error estimates of the Royston (Roy), Srivastava-Hui (S-H), and Doornik-Hansen test (D-H) are found very close to the nominal alpha level (0.05). Generally, the type I error rates of these tests varied from 4.9 to 5.5% regardless with sample size (n) and number of variables (p). On the other hand, the type I error rates of Mardia's Skewness (M-S), Mardia's Kent (M-K) and Henze-Zirkler Test (H-Z) varied based on their sample size and number of variables. Among these three tests, the closest estimations to the nominal alpha level have been observed when H-Z test was applied. For H-Z test, the type I error rates were found to be between 2.3 and 5.5%. The most reliable results have been obtained when sample size of 30≤n≤50 regardless number of variables while the most deviated results were noticed when sample size of n≤20 in case of H-Z test. The type I error estimates of M-S test have generally been found around 6.00% (varied from 6.0 to 7.2%). The most deviated results, among the M-S, M-K and H-Z, were obtained when M-K test was used under sample sizes of n≥50. The M-K test showed reliable results especially when sample size was 30.

3.2 Results of Test Power

The estimations of test power for six different tests have been given in Tables 2, 3 and 4. The test power estimates of all tests have been found lowest in range when samples were taken from multivariate t-distribution with 10 d.f. (Table 2). The maximum test power value has been obtained when D-H (67.8 %) and M-K (63.2%) tests were applied under n=100 and p=5. None of these tests reached enough to test power value of 80.00% under such experimental circumstances. In contrast, the test power estimates under multivariate t-distribution, all six tests performed highly test power for skewed distributions namely Multivariate Cauchy (Table 3) and Multivariate Lognormal (Table 4) distributions. For both of the multivariate Cauchy as well as multivariate Lognormal distributions, all estimations of each test power have been recorded as 100.00% in entire combinations of *n* and *p* exceptionally in a few cases when *n* = 20. The lowest test power was appeared when *n* = 20 and *p* = 2 for M-K test that has been achieved as 87.2%. As far as the Roy and D-H tests are concerned, both of them have always attained to 100% test power regardless of sample size and number of variable.

3.3 Results of Multidimensional Scaling Technique

Multidimensional Scaling Technique (MDS) has been applied aimed to classify the said tests in terms of their performances (Type I error rate and test power) under all considered experimental conditions. In this way, it would be possible to determine the tests that revealed similar performance. The results related to the test have been shown in Figs. 1, 2, 3, 4, 5 and 6. Multidimensional scaling is an exploratory technique used to visualize proximities (a proximity is a number that indicates how similar or how different two objects or variables) in a low dimensional space. MDS provides a researcher to uncover the hidden structure or relations among the variables. Each object is represented by a point in a multidimensional space. Two similar objects are represented by two points that are close to each other, while two different or dissimilar objects are represented by two points that are apart from each other [15].

MDS analysis performed to classify these tests with respect to type I error estimates under different sample size and number of variable combinations (Fig. 1). Two different goodness-of-fit criteria namely R^2 and stress coefficient have been used to determine the suitability of MDS technique to assess the type I error estimates of all six tests simultaneously. R^2 and stress coefficient values (0.994 and 0.029) indicated that MDS technique was one of a good choice to evaluate the performances of these tests.

Similar and different tests with respect to their Type I error estimates under all sample size and number of variable combinations are given in Fig. 1. As the Fig. 1 shows that the D-H, S-H, and Roy tests are placed in the same group that means these tests produced very similar results in terms of retaining type I error rate at the nominal alpha level (0.05) or the performances of these tests are quite similar to each other. On the other hand, the M-S, H-Z, and M-K tests are located in different places. Consequently, the type I error estimates of M-S, H-Z, and M-K tests were found different when compared them to the D-H, S-H, and Roy tests. It is also possible to conclude that the performances of M-K are obviously different from the D-H, S-H, and Roy tests.

In order to determine those factors that affect Type I error estimates of six tests are given in Fig. 2. The Results of MDS showed that the type I error estimates have generally been affected by sample size and number of variables (Fig. 2). Differences among the tests were quite obvious especially when n=20 and n=30. As sample size is increased then the effect of number of variables on type I error estimates decreased to ignorable level.

Figs. 3, 4 and 5 have been established by using MDS technique aimed to determine similar tests in terms of test power estimates when samples were taken from multivariate Cauchy, multivariate lognormal, and multivariate t-distribution with 10 d.f., respectively.

Table 1. Type I error rates when samples are taken from multivariate normal distribution

	n=20				**n=30**				**n=50**				**n=100**			
	p=2	**p=3**	**p=4**	**p=5**	**p=2**	**p=3**	**p=4**	**p=5**	**p=2**	**p=3**	**p=4**	**p=5**	**p=2**	**p=3**	**p=4**	**p=5**
M-S	6.3	6.0	6.3	6.1	6.2	6.8	6.9	7.0	6.2	6.7	6.9	7.2	6.2	6.3	6.3	6.6
M-K	3.9	4.5	3.8	4.0	4.8	5.7	5.6	5.2	6.2	7.1	8.0	8.2	6.3	8.5	9.7	10.5
D-H	**5.0**	**5.2**	**5.2**	**4.9**	**5.1**	**5.2**	**5.0**	**5.1**	**5.4**	**5.2**	**5.3**	**5.4**	**5.1**	**5.3**	**5.5**	**5.4**
S-H	**5.1**	**5.1**	**5.1**	**5.2**	**5.1**	**5.1**	**5.0**	**5.1**	**5.3**	**5.3**	**5.3**	**5.4**	**5.3**	**5.4**	**5.4**	**5.5**
H-Z	3.5	2.9	2.7	2.3	4.9	4.9	5.0	5.1	5.5	5.3	5.2	5.1	4.4	4.2	4.0	3.9
Roy	**5.2**	**5.3**	**5.1**	**5.0**	**5.1**	**5.3**	**5.2**	**5.3**	**5.4**	**5.4**	**5.5**	**5.2**	**5.2**	**5.3**	**5.1**	**5.1**

Table 2. Test power of the tests when samples are taken from multivariate t- distribution with 10 d.f

	n=20				**n=30**				**n=50**				**n=100**			
	p=2	**p=3**	**p=4**	**p=5**	**p=2**	**p=3**	**p=4**	**p=5**	**p=2**	**p=3**	**p=4**	**p=5**	**p=2**	**p=3**	**p=4**	**p=5**
M-S	15.9	17.1	16.8	16.7	19.5	21.6	22.5	23.9	23.2	26.7	30.2	32.0	29.5	33.4	38.5	41.7
M-K	**12.0**	**12.0**	**10.6**	**8.3**	**18.4**	**21.9**	**21.9**	**21.2**	**28.3**	**33.7**	**37.3**	**38.7**	**44.3**	**52.2**	**59.6**	**63.2**
D-H	**15.1**	**16.9**	**17.8**	**19.1**	**20.7**	**24.4**	**25.8**	**28.5**	**28.9**	**33.9**	**39.4**	**43.1**	**44.0**	**53.4**	**61.9**	**67.8**
S-H	11.3	11.5	11.0	10.6	13.3	12.8	12.6	12.0	13.8	13.0	13.2	13.1	12.2	11.7	11.1	10.0
H-Z	8.3	8.1	7.3	7.9	10.3	9.8	9.1	8.6	11.6	11.3	11.7	10.4	15.4	16.0	16.2	15.1
Roy	13.0	15.1	15.9	18.3	15.7	18.4	19.1	21.8	16.6	19.1	22.0	24.6	34.5	41.4	48.8	53.3

Table 3. Test power of the tests when samples are taken from multivariate Cauchy distribution

	n=20				**n=30**				**n=50**				**n=100**			
	p=2	**p=3**	**p=4**	**p=5**	**p=2**	**p=3**	**p=4**	**p=5**	**p=2**	**p=3**	**p=4**	**p=5**	**p=2**	**p=3**	**p=4**	**p=5**
M-S	96.9	98.5	99.9	99.8	100	100	100	100	100	100	100	100	100	100	100	100
M-K	96.6	99.2	99.5	99.7	100	100	100	100	100	100	100	100	100	100	100	100
D-H	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
S-H	97.1	99.2	100	100	100	100	100	100	100	100	100	100	100	100	100	100
H-Z	97.7	98.8	99.4	99.9	100	100	100	100	100	100	100	100	100	100	100	100
Roy	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Table 4. Test power of the tests when samples are taken from multivariate lognormal distribution

	n=20				**n=30**				**n=50**				**n=100**			
	p=2	**p=3**	**p=4**	**p=5**	**p=2**	**p=3**	**p=4**	**p=5**	**p=2**	**p=3**	**p=4**	**p=5**	**p=2**	**p=3**	**p=4**	**p=5**
M-S	95.6	98.2	98.3	98.5	99.4	100	100	100	100	100	100	100	100	100	100	100
M-K	87.2	91.2	93.1	93.2	98.2	100	100	100	100	100	100	100	100	100	100	100
D-H	100	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
S-H	**94.1**	**96.2**	**95.7**	**96.3**	**99.5**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
H-Z	**96.8**	97.8	98.4	99.1	100	100	100	100	100	100	100	100	100	100	100	100
Roy	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**

All tests were located in the same group when samples have been taken from multivariate Cauchy and multivariate lognormal shown in Figs. 3 and 4. It indicated that when samples were taken from distributions with high skewness and kurtosis (multivariate lognormal and multivariate Cauchy distributions) then the test power estimates of all tests were found highly and similar to each other. Therefore, it is possible to conclude that the test power of these tests do not affect from distribution shapes as long as samples are taken from distributions with high skewness and kurtosis values.

When samples were taken from multivariate t-distribution with 10 d.f. (Fig. 3) then it was observed that the tests have been assigned to three different groups. Such as, the M-K and D-H tests have been placed to first group, H-Z and S-H tests found in second group while the Roy and M-S were located in the third group. The tests that were situated in the first group, among these three groups, had high performance when compared to the tests that were located in the second and third groups.

Factors that affect test power estimates have also been determined by using MDS and the results were given in Fig. 6. According to the results, the main factors that affect the test power estimates were known as skewness and kurtosis of distribution (distribution shape) where the samples have been taken from (Fig. 6). Cauchy and lognormal distributions were taken place in the same group while multivariate t-distribution with 10 d.f found in a different group as shown in Fig. 6. That is why, it is possible to conclude that it is expected to obtain quite similar test power estimates as long as samples were taken from distributions with high skewness and kurtosis.

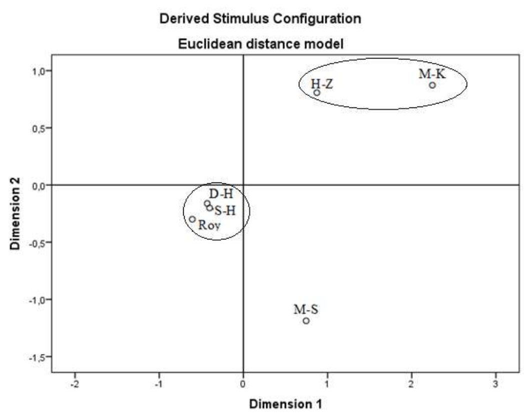

Fig. 1. MDS result for assessing type I error estimates

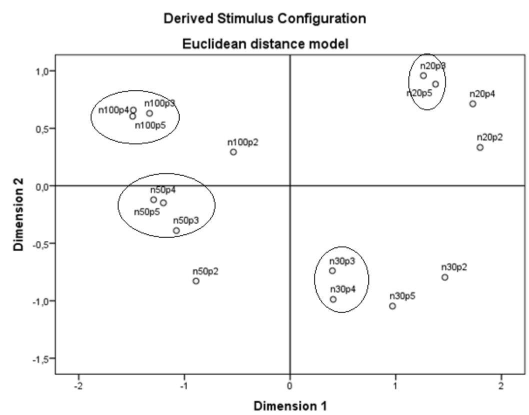

Fig. 2. MDS result for determining factors that affect type I error estimates

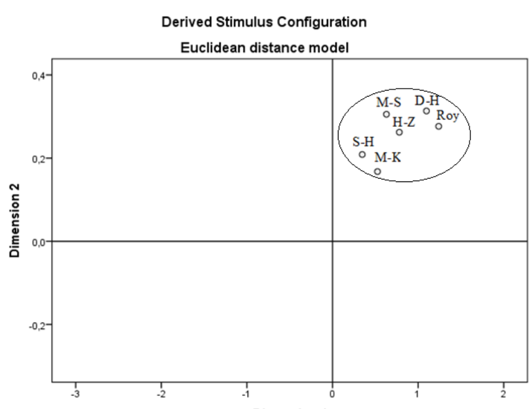

Fig. 3. MDS result for assessing test power estimates when sample are taken from multivariate lognormal distribution

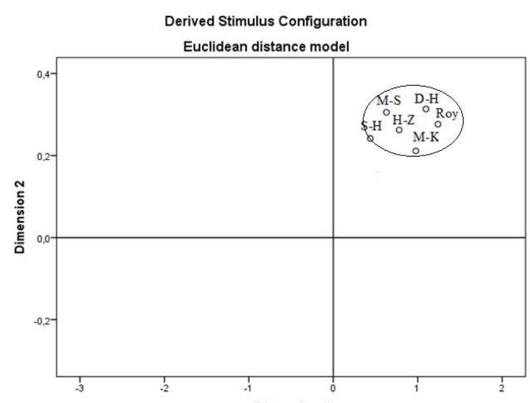

Fig. 4. MDS result for assessing test power estimates when sample are taken from multivariate Cauchy distribution

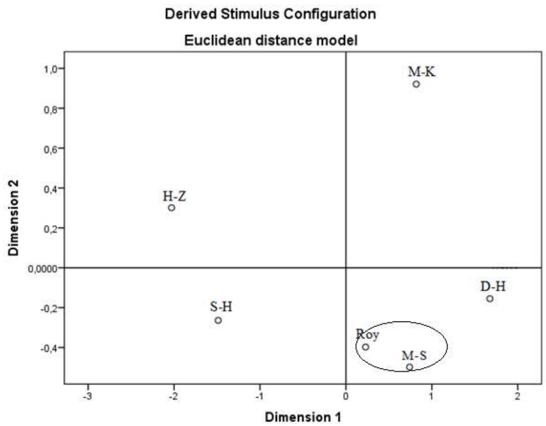

Fig. 5. MDS result for assessing test power estimates when sample are taken from multivariate t- distribution with 10 d.f

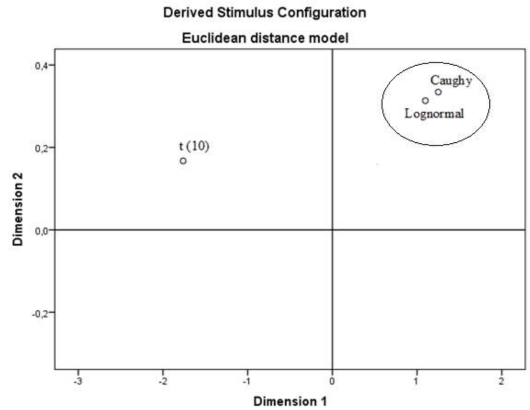

Fig. 6. MDS results for determining factors that affect test power estimates

4. DISCUSSION AND CONCLUSION

It is known very well that the performance of many multivariate methods is affected by deviation from normality [1,2,16]. However, the assumption of multivariate normality has often been ignored [17]. This situation may lead to the achievement of misleading results. That is why, the answer of such question is important here: which test(s) should be applied for testing the MVN assumption? It is possible to answer this question by determining the tests having similar performance by considering both of the type I error and the test power estimates of the said tests. Firstly, a comprehensive simulation study has been carried out for obtaining the empirical type I error and test power estimates of the Royston (Roy), Srivastava-Hui (S-H), and Doornik-Hansen test (D-H), Mardia's Skewness (M-S), Mardia's Kent (M-K) and Henze-Zirkler (H-Z) tests in this regard. Then, Multidimensional Scaling (MDS) technique has been applied to classify the tests those have had similar performance and the factors affected the performances of those tests. The obtained results of this study showed that the Roy, S-H, and D-H tests are generally more appropriate tests whether to use in test of MVN assumption or not in respect to regardless of sample size, number of variables and distribution shape. On the other hand, the performances of the rest tests changed depending on experimental conditions. Thus, we suggested the application of the Roy, S-H, or D-H tests for testing MVN assumption before performing an inferential multivariate procedure like MANOVA, Discriminant Analysis, and/or Multivariate Multiple Regression for analyzing the multivariate data assume to MVN. Doornik and Hansen [12] compared their proposed method with Mardia [18], Royston extension of the Shapiro and Wilk [19] test. They reported that their test had a good performance in terms of retaining type I error rate at the nominal level and good test power values. Naczk [20] reported that the Henze-Zirkler test was the only test that could be recommended for the assessment of multivariate normality. Mecklin and Mundfrom [1] reported that there was no single test found to be the most powerful under all circumstances during the simulation studies when comparing the performance of 13 MVN tests depending on 10000 simulations. They recommend Henze- Zirkler test as a formal test for testing the MVN.

Farrell et al. [2] conducted a simulation study to compare the size and power of the Royston [10], Doornik and Hansen [12], and Henze and Zirkler [8] test. They reported that the Royston's test produced the best results regarding empirical Type I error rates, which ranged between 4.54% and 5.26% over all combinations of n and p. The estimates for the D-H test were also extremely good in all cases. They also reported that the Henze and Zirkler (H-Z) test generally possesses good power across the alternative distributions investigated particularly for $n >= 75$ while it is not satisfactory when detecting the reason(s) of separation from MVN.

While comparing the simulation results of our research work with those of other studies where we noticed some important differences and it is assumed that these differences may be resulted due to the usage of different data generation routines, differences in experimental conditions and number of running simulation. As a results, the Royston (Roy), Srivastava-Hui (S-H), and Doornik-Hansen test (D-H) tests, based on our experimental conditions, were recommended for assessing MVN assumptions in general.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Mecklin CJ, Mundfrom DJ. A Monte Carlo comparison of the type I and type II error rates of tests of multivariate normality. J Stat Comput Sim. 2005;75(2):93-107.
2. Farrel PJ, Salibian-Barrera M, Naczk K. On Tests for multivariate normality and associated simulation studies. J Stat Comput Sim. 2007;77(12):1065−1080.
3. Korkmaz S, Goksuluk D, Zararsiz G. MVN: An R Package for Assessing Multivariate Normality. 2015;6(2):151-162.
4. Hopkins JW, Clay PPF. Some empirical distributions of bivariate T2 and homoscredasticity criterion M under unequal variance and leptokurtosis. J Am Stat Assoc. 1963;58(304):1048-1053.
5. Mardia KV. Assessment of multinormality and the robustness of Hotelling's T2 test. Appl Statist. 1975;24(2):163-171.
6. Conover WJ, Iman RL. The rank transformation as a method of discrimination with some examples. Commun Stat Theor. 1980;9(5):465-487.
7. Royston JP. Some techniques for assessing multivariate normality based on the Shapiro-Wilk W. Appl Statist. 1983; 32(2):121-133.
8. Henze N, Zirkler B. A class of invariant consistent tests for multivariate normality. Commun Stat Theor. 1990;19(10):3595-3617.
9. Horswell RL, Looney SW. A comparison of tests for multivariate normality that are based on measures of multivariate skewness and kurtosis. J Stat Comput Sim. 1992;42(1-2):21-38.
10. Royston JP. Approximating the Shapiro-Wilk W-Test for non-normality. Stat Comput. 1992;2(3):117-119.
11. Romeu JL, Ozturk A. A comparative study of goodness of fit tests for multivariate normality. J Multivariate Anal. 1993;46(2): 309-334.
12. Doornik JA, Hansen D. An omnibus test for univariate and multivariate normality. Oxford Bulletin of Economics and Statistics. 2008;70:927–939.
13. Mecklin CJ. A comparison of the power of classical and newer tests of multivariate normality. Doctoral Dissertation, University of Northern Colorado; 2000.
14. Srivastava MS, Hui TK. On assessing multivariate normality based on the Shapiro Wilk W statistic. Stat Probabil Lett. 1987;5(1):15-18.
15. Kruskal JB, Wish M. Multidimensional Scaling. Beverly Hills and London: Sage Publications; 1978.
16. Looney SW. How to use tests for univariate normality to assess multivariate normality. Am Stat. 1995;49(1):64-70.
17. Baxter MJ. Testing multivariate normality, with applications to lead isotope data analysis in archaeology. Bar International Series. 1997;750:107-107.
18. Mardia KV. Measures of multivariate skewness and kurtosis with applications. Biometrika. 1970;57(3):519-530.
19. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika. 1965;52(3-4):591-611.
20. Naczk K. Assessing tests for multivariate normality. MSc Thesis, School of Mathematics and Statistics, Carleton University, Ottawa, Ontario; 2004.

Peer-review history:
The peer review history for this paper can be accessed here:
http://sciencedomain.org/review-history/14595