

Original Research Article

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

Abstract

This simulation study has been conducted to evaluate the performances of six different multivariate normality tests under different experimental conditions. Firstly, empirical Type I error rate and test power of Mardia's Skewness, Royston, Srivastava-Hui, Doornik-Hansen Test, Henze-Zirkler, and Mardia's Kent have been obtained. Then, Multidimensional Scaling (MDS) technique has been applied to classify the tests those have had similar performance and the factors those affected the performances of the above mentioned tests. At the end of Multidimensional Scaling analyses, it has been observed that the Royston, Srivastava-Hui and Doornik-Hansen tests showed similar performance, and the performances of these tests were obviously different than that of the others in general.

Key words: 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 by Mecklin and Mundfrom, 2005; Farrell et al., 2006. 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 (Mardia, 1975; Royston, 1983, Henze and Zirkler, 1990; Horswell and Looney, 1992; Royston, 1992; Romeu and Ozturk, 1993; Doornik and Hansen, 1994; Mecklin, 2000; Mecklin and Mundfrom, 2005; Farrell et al., 2006).

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 (Mardia, 1975; Royston, 1983; Srivastava and Hui, 1987; Henze and Zirkler, 1990; Royston, 1992; Romeu and Ozturk, 1993; Doornik and Hansen, 1994; Mecklin, 2000; Mecklin and Mundfrom, 2005; Farrell et al., 2006). 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. Material 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 $\text{Rho}=0.60$. 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. The above mentioned tests have been classified based on their performance by using Multidimensional Scaling Technique (MDS).

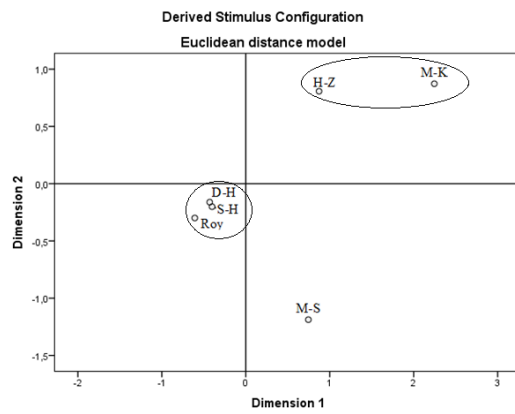
3. Results

Multidimensional Scaling Technique (MDS) has been applied aimed to classify the said tests in terms of their performances 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 Figure 1, 2, 3, 4, 5, and 6. MDS analysis performed to classify these tests with respect to type I error estimates under different sample size and number of variable combinations (Figure 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. As the Figure 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 Figure 2. The Results of MDS showed that the type I error estimates have generally been affected by sample size and number of variables (Figure 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. Figure 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. All tests were located in the same group when samples have been taken from multivariate Cauchy and multivariate lognormal shown in Figure 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. (Figure 5) 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 Figure 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 (Figure 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 Figure 6. That is why, it is possible to conclude that it is expected to obtain quite similar

99
100

test power estimates as long as samples were taken from distributions with high skewness and kurtosis.

Figure 1. MDS results for assessing type I error estimates



101

Figure 2. MDS results for determining factors that affect type I error estimates

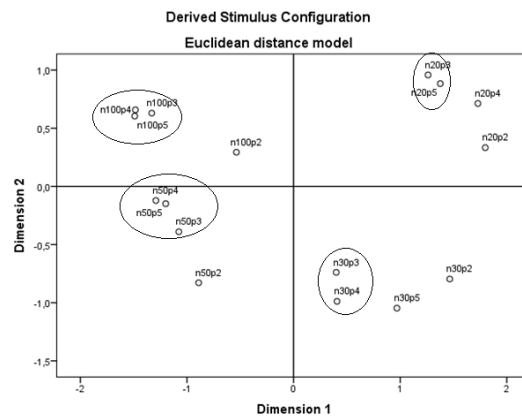
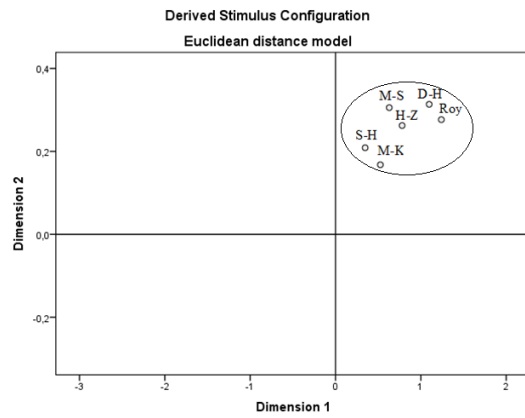


Figure 3. MDS results for assessing test power estimates when samples are taken from multivariate lognormal distribution



102

Figure 4. MDS results for assessing test power estimates when samples are taken from multivariate Cauchy distribution

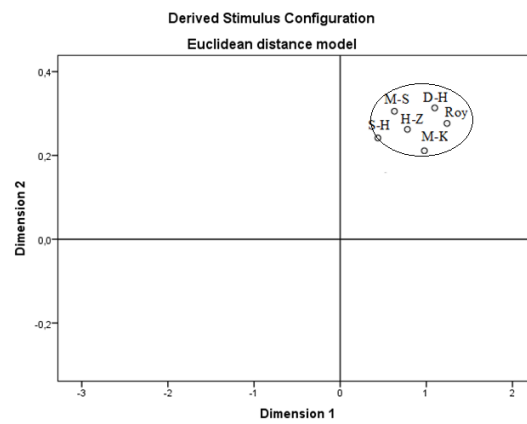
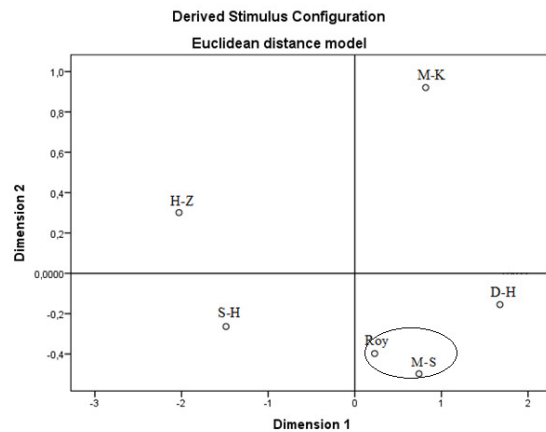
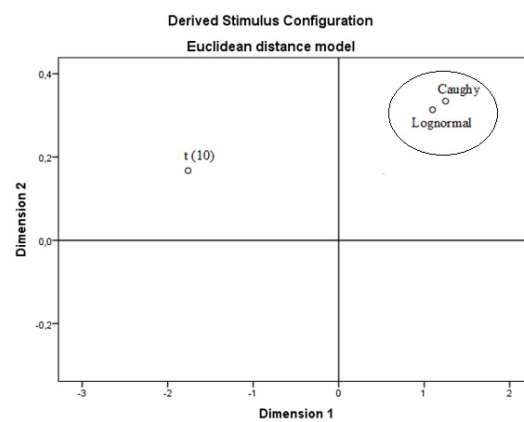


Figure 5. MDS results for assessing test power estimates when samples are taken from multivariate t-distribution with 10 d.f.



103

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



104

105 4. Discussion and Conclusion

106 It is known very well that the performance of many multivariate methods is affected by deviation
 107 from normality (Looney, 1995; Mecklin and Muldfrem, 2005, Farrell et al., 2006). However, the
 108 assumption of multivariate normality has often been ignored (Baxter, 1997). This situation may lead
 109 to the achievement of misleading results. That is why, the answer of such question is important
 110 here: which test(s) should be applied for testing the MVN assumption? It is possible to answer this
 111 question by determining the tests having similar performance by considering both of the type I error
 112 and the test power estimates of the said tests. Firstly, a comprehensive simulation study has been
 113 carried out for obtaining the empirical type I error and test power estimates of the Royston (Roy),
 114 Srivastava-Hui (S-H), and Doornik-Hansen test (D-H), Mardia's Skewness (M-S), Mardia's Kent
 115 (M-K) and Henze-Zirkler (H-Z) tests in this regard. Then, Multidimensional Scaling (MDS)
 116 technique has been applied to classify the tests those have had similar performance and the factors
 117 affected the performances of those tests. The obtained results of this study showed that the Roy, S-
 118 H, and D-H tests are generally more appropriate tests whether to use in test of MVN assumption or
 119 not in respect to regardless of sample size, number of variables and distribution shape. On the other
 120 hand, the performances of the rest tests changed depending on experimental conditions. Thus, we
 121 suggested the application of the Roy, S-H, or D-H tests for testing MVN assumption before
 122 performing an inferential multivariate procedure like MANOVA, Discriminant Analysis, and/or
 123 Multivariate Multiple Regression for analyzing the multivariate data assume to MVN. Doornik and
 124 Hansen (1994) compared their proposed method with Mardia (1970), Royston extension of the
 125 Shapiro and Wilk (1965) test. They reported that their test had a good performance in terms of
 126 retaining type I error rate at the nominal level and good test power values. Naczka (2004) reported
 127 that the Henze-Zirkler test was the only test that could be recommended for the assessment of
 128 multivariate normality. Mecklin and Mundfrom (2005) reported that there was no single test found
 129 to be the most powerful under all circumstances during the simulation studies when comparing the
 130 performance of 13 MVN tests depending on 10000 simulations. They recommend Henze- Zirkler
 131 test as a formal test for testing the MVN.

132 Farrell et al (2006) conducted a simulation study to compare the size and power of the Royston
 133 (1992), Doornik and Hansen (1994), and Henze and Zirkler (1990) test. They reported that the
 134 Royston's test produced the best results regarding empirical Type I error rates, which ranged
 135 between 4.54% and 5.26% over all combinations of n and p . The estimates for the D-H test were
 136 also extremely good in all cases. They also reported that the Henze and Zirkler (H-Z) test generally
 137 possesses good power across the alternative distributions investigated particularly for $n \geq 75$ while
 138 it is not satisfactory when detecting the reason(s) of separation from MVN.

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

145

146 5. References

147

- 148 Baxter, M. J. (1999). Testing Multivariate Normality, with Applications to Lead Isotope Data
 149 Analysis in Archaeology. *Bar International Series*, 750, 107-107.
 150 Doornik, J.A., Hansen, D.,(1994). An omnibus test for univariate and multivariate normality. Working
 151 Paper, Nuffield College, Oxford.
 152
 153 Farrel P.J., Salibian-Barrera, Naczka K. (2007). On Tests for multivariate normality and associated
 154 simulation studies. *J Stat.Comput.Sim.*, 77, 1065–1080.
 155 Henze, N.,Zirkler, B., (1990). A class of invariant consistent tests for multivariate normality.

- 156 *Commun. Stat.Theor.*, 19, 3595-3617
- 157 Horswell, R.L.,Looney, S.W., (1992). A comparison of tests for multivariate normality that are based
- 158 on measures of multivariate skewness and kurtosis. *J Stat.Comput.Sim.*, 42, 21-38.
- 159 Looney, S.W.(1995). How to use tests for univariate normality to assess multivariate normality. *Am.*
- 160 *Stat.*, 49, 64-70.
- 161 Mardia, K.V.,(1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*,
- 162 57, 519-530.
- 163 Mardia, K.V.,(1975). Assessment of multinormality and the robustness of Hotelling's T2 test. *Appl.*
- 164 *Statist.*, 24, 163-171.
- 165 Mecklin, C.J.(2000). A comparison of the power of classical and new tests of multivariate normality.
- 166 Doctoral Dissertation, University of Northern Colorado.
- 167 Mecklin, C.J., Mundfrom, D.J. (2005). A Monte Carlo comparison of the Type I and Type II error
- 168 rates of tests of multivariate normality. *J Stat.Comput.Sim.*, 75, 93 - 107.
- 169 Naczki, K.,(2004). Assessing tests for multivariate normality. MSc Thesis, School of Mathematics and
- 170 Statistics, Carleton University, Ottawa, Ontario.
- 171 Royston, J.P.,(1983). Some techniques for assessing multivariate normality based on the Shapiro-
- 172 Wilk W. *Appl.Statist.*, 32, 121-133.
- 173 Royston, J.P.,(1992). Approximating the Shapiro-Wilk W-Test for non-normality. *Stat. Comput.*, 2,
- 174 117-119.
- 175 Romeu, J.L.,Ozturk, A., (1993). A comparative study of goodness of fit tests for multivariate
- 176 normality. *J Multivariate Anal.*, 46, 309-334.
- 177 Srivastava, M.S., Hui, T.K., (1987). On assessing multivariate normality based on the Shapiro Wilk
- 178 W statistic. *Stat. Probabil.Lett.*, 5, 15-18.