On Using Asymptotic Critical Values in Testing for Multivariate Normality

Christopher J. Mecklin[1*] and Daniel J. Mundfrom[2]

[1]*Department of Mathematics and Statistics, Murray State University;* [2]*Department of Applied Statistics and Research Methods, University of Northern Colorado.*

Abstract: Many multivariate statistical methods call upon the assumption of multivariate normality. However, many researchers fail to test this assumption. This omission could be due either to ignorance of the existence of tests of multivariate normality or confusion about which test to use. Although at least 50 tests of multivariate normality exist, relatively little is known about the power of these procedures. The purpose of this study was to examine the power (in small sample situations) of 8 promising tests of multivariate normality with a Monte Carlo study. Ten thousand data sets were generated from several multivariate distributions. The test statistic for each procedure was calculated and compared with the appropriate asymptotic critical value. The number of rejections of the null hypothesis of multivariate normality was tabled for each situation. No single test was found to be the most powerful in all situations. The use of the Henze-Zirkler test is recommended as a formal test of multivariate normality. Supplementary procedures such as Mardia's skewness and kurtosis measures and the chi-square plot are also recommended for diagnosing possible deviations from normality.

Keywords: Power comparison; Mardia's skewness and kurtosis; Empirical characteristic function; Goodness-of-fit

1. INTRODUCTION

It is well known that many multivariate statistical procedures, including MANOVA, discriminant analysis, and canonical correlation, call upon the assumption of multivariate normality (MVN). Looney (1995) found that the performance of these procedures is affected to various degrees by certain deviations from normality. Mardia et al. (1979) and DeCarlo (1997) stated that hypothesis tests involving mean vectors are more sensitive to the effects of skewness, while tests involving variance-covariance matrices are more sensitive to kurtosis. However, the assumption of multivariate normality often goes untested. Horswell (1990) declared that tests of MVN are "largely academic curiosities, seldom used by practicing statisticians." Looney (1995) gave several reasons for the failures to check for MVN, including ignorance of the existence of statistical tests for MVN, lack of readily available software to conduct the tests, and the reluctance of practitioners to use a procedure when little is known about the power of the procedure.

There is no shortage of procedures to test the goodness-of-fit of a data set to the multivariate normal distribution. At least 50 different procedures have been proposed for this problem. Because there are countless possible deviations from normality, Andrews et al. (1973) concluded that multiple approaches for testing MVN

[1]E-mail:[1]christopher.mecklin@murraystate.edu; [2]djmundf@unco.edu

*Corresponding author: Department of Mathematics and Statistics, Murray State University, Murray, KY 42071, USA

would be needed. Unfortunately, different conclusions about the MVN of a data set can be reached by different procedures. As an example, consider the famous *Iris setosa* data set originally considered by Fisher and available in many multivariate textbooks such as Johnson and Wichern (1992). Multivariate normality was not rejected at the $\alpha = 0.05$ level using the procedures of Mardia (1970),Koziol (1982), Tsai and Koziol (1988), Mardia and Kent (1991), or Koziol (1993). On the other hand, Small (1980), Royston (1983), Baringhaus and Henze (1988), and Romeu and Ozturk (1996) rejected MVN at $\alpha = 0.05$ for the Iris setosa data set.

The main goal of this investigation was to identify those procedures for testing MVN that are effective against a wide range of non-normal alternatives. The procedure(s) thus identified as effective could then be employed by a researcher even and especially when the true distribution of the population is not known *a priori*. More specifically, 8 different tests of multivariate normality that utilize asymptotic critical values (i.e. the test statistic has an asymptotic null distribution)were compared in a Monte Carlo simulation study against 25 different multivariate distributions. These distributions ranged from the multivariate normal to severe deviations from normality.

2. BACKGROUND

There is no shortage of proposed methods for assessing multivariate normality. A current review of the literature revealed that at least 50 procedures exist to test for MVN. Several earlier reviews, such as Looney (1995), Andrews et al. (1973), Gnanadesikan (1977), Mardia (1980), and Koziol (1986) of the literature exist. Explanations of some of the simpler methods can be found in most multivariate textbooks, such as Johnson and Wichern (1992) or Rencher (1995). However, Rencher (1995) commented that since multivariate normality is not as straightforward as univariate normality, the state of the art is not as refined.

When compared to the amount of research conducted in developing tests of MVN, a relative paucity of work exists in evaluating the quality, in terms of Type I and Type II error rates and power, of these procedures. Ward (1988), Horswell and Looney (1992), Romeu and Ozturk (1993), Young et al. (1995), and Bogdan (1999) have conducted studies comparing the power of various tests of multivariate normality. Most of these studies were deliberately restricted in scope to a limited category of tests.

Much of the field of multivariate statistics consists of extending univariate methods to the general case. Testing for goodness-of-fit to a hypothesized MVN distribution is no exception. Most of the available tests of MVN are extensions of simpler procedures for testing univariate normality. Thus, a large proportion of available tests of MVN are based on normal probability plots, measures of skewness and kurtosis, or goodness-of-fit procedures. Unfortunately, Bogdan (1999) and Koziol (1983) found that few of these tests are formal, in the sense that both the null distribution of the test statistic has been found and the consistency of the test has been established. In the words of Baringhaus and Henze (1988) and Csörgö (1989), there are few “genuine” tests of MVN.

A seminal paper of Mardia (1970) introduced multivariate measures of skewness and kurtosis, denoted by $\beta_{1,p}$ and $\beta_{2,p}$, respectively. For the multivariate normal distribution, $\beta_{1,p} = 0$ and $\beta_{2,p} = p(p+2)$. It was determined that a function of the multivariate skewness is asymptotically distributed as a chi-square random variable with $\frac{p(p+1)(p+2)}{6}$ degrees of freedom and a function of the multivariate kurtosis is asymptotically distributed as a standard normal random variable. These asymptotic distributions were exploited to develop two tests of multivariate normality.

Mardia's procedures, particularly the test based on multivariate kurtosis, are among the most commonly used tests of MVN, especially within the structural equation modeling community. Previous research by Ward (1988), Horswell and Looney (1992), Romeu and Ozturk (1993), and Bogdan (1999) has indicated that these procedures are among the best available tests of MVN. It is inconceivable that any comprehensive study of the performance of MVN tests would not include Mardia's skewness and kurtosis. Thus, both were considered in this study.

Efforts have been made to combine elements of multivariate skewness and kurtosis into a single 'omnibus' test statisitic. Mardia and Foster (1983) derived six possible test statistics. A statistic denoted S_W^2, which used the Wilson-Hilferty approximation, had an asymptotic chi-square distribution with 2 degrees of freedom. Ward (1988) and Horswell and Looney (1992) found S_W^2 to lack power. An alternative statistic C_W^2, which included the covariance between multivariate skewness and kurtosis, was not previously evaluated. A more recently developed omnibus test, by Mardia and Kent (1991), was derived using Rao scores. Both C_W^2 and the Mardia-Kent test were considered.

Empirical distribution function tests, such as the Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling, can be used to test for univariate normality. A multitude of researchers has extended these procedures and other goodness-of-fit procedures to the general multivariate case in order to develop a test of MVN. Although many promising tests of MVN are included in this class, Rencher (1995) criticized this approach as unrealistic due to the inherent sparseness of multivariate data.

The Shapiro-Wilk test, Shapiro and Wilk (1965), is generally considered to be an excellent test of univariate normality. It is only natural to extend it to the multivariate case, as done by both Royston (1983) and Mudholkar et al. (1995). A recently developed and less well known goodness-of-fit procedure is the Q_n procedure of Ozturk and Dudewicz (1992). Romeu and Ozturk (1993) developed two versions of the Q_n procedure to test for MVN. Their own power study indicated that the Cholesky root version of this test was one of the most powerful procedures available.

All of the above procedures, from the first three classes of tests of MVN, have been criticized by Baringhaus and Henze (1988), Bogdan (1999), Csörgö (1989), and Henze and Zirkler (1990) for lacking the property of consistency. This same criticism is applicable to the univariate tests of normality that the MVN tests were based upon. A recent breakthrough in addressing the issue of consistency is the development of a class of procedures that are consistent, typically by using the empirical characteristic function.

The characteristic function of some continuous random variable X is

$$\phi(x) = E(e^{itx}) \tag{2.3}$$

which, as pointed out in Kendall and Stuart (1977), is actually a special case of a Fourier transform that involves distribution functions. The empirical characteristic function is defined as

$$\phi_n(t) = \frac{1}{n}\sum_{j=1}^{n} e^{itX_j}, \tag{2.4}$$

where $X_j, j = 1, \cdots, n$ is a random sample from some distribution $F(x)$. Epps and Pulley (1983) proposed a test of the composite hypothesis of univariate normality based upon the statistic

$$T = \int_{-\infty}^{\infty} |\phi_n(t) - \hat{\phi_O}(t)|^2 dG(t), \tag{2.5}$$

where $\hat{\phi_O}(t)$ is the characteristic function of the normal distribution which is estimated using the sample mean and variance and $G(t)$ is a weight function. Epps and Pulley (1983) also obtained a closed form for their test statistics and found that their test was quite competitive with the Shapiro-Wilk procedure. They did not, however, prove the consistency of this test.

The first attempt to extend this idea to testing MVN was by Csörgö (1986). Baringhaus and Henze (1988) extended the Epps-Pulley test to the multivariate case and obtained an asymptotic null distribution for their test statistic. Csörgö (1989) proved the consistency of the Epps-Pulley and Baringhaus-Henze tests. Henze and Zirkler (1990) extended the tests of Baringhaus-Henze and Epps-Pulley into their current form. The Henze-Zirkler test statistic is based on a non-negative functional that measures the distance between two distributions, the hypothesized distribution (MVN) and the observed distribution. For consistency to hold, this functional must equal zero if and only if the observed data is MVN. The non-negative functional considered is:

$$D_\beta(P, Q) = \int |\hat{P}(t) - \hat{Q}(t)|^2 \varphi_\beta(t) dt, \tag{2.6}$$

where $\hat{P}(t)$ is the characteristic function of the MVN distribution, $\hat{Q}(t)$ is the empirical characteristic function, $\varphi_\beta(t)$ is a kernel function which was chosen by Henze and Zirkler (1990) to be $N_p(0, \beta^2 I_p)$, and β is a smoothing parameter. Extensive derivations, found in Henze and Zirkler (1990), yielded a closed form for the Henze-Zirkler statistic which had an approximate lognormal distribution.

The proof of the consistency of the Henze-Zirkler test statistics followed directly from Csörgö (1989). Since this procedure had not been considered in a previous power comparison study and the consistency implies the strong possibility that this test will be very competitive against a wide range of alternatives, it was included in this research.

3. METHODS OF SIMULATION STUDY

The purpose of this study was to compare the Type I and Type II error rates (and hence the power) of a variety of tests for MVN that use asymptotic critical values. To accomplish this, a Monte Carlo simulation study was designed. In the study, 10000 data sets with combinations of sample size $n = 25, 50, 100$ and dimension $p = 2, 3, 4, 5$ were generated from 25 different multivariate distributions, ranging from the multivariate normal to severe departures from normality. With three different sample sizes, four different dimensions, and 25 different distributions, a total of 300 situations were simulated. The limitations on sample size and dimension were imposed to keep the amount of computing time reasonable and to consider sample sizes that are minimal for multivariate analysis. These small sample sizes

Table 1. Tests of Multivariate Normality

Test	*Iris setosa*
Mardia's Skewness	Do not reject
Mardia's Kurtosis	Do not reject
Mardia-Foster	Reject
Royston	Reject
Henze-Zirkler	Do not reject
Mardia-Kent	Do not reject
Romeu-Ozturk	Reject
Mudholkar-Srivastava-Lin	Do not reject

are likely to be the situation where the assumption of multivariate normality is most critical to the researcher. The test statistics were calculated for all 10000 data sets and compared to the appropriate critical value in order to estimate the proportion of rejections of MVN for each test in each situation. The standard level of significance of $\alpha = 0.05$ was used.

A total of 8 different tests of multivariate normality were studied in this investigation. A list of these procedures and the results of this procedure when applied on the *Iris setosa* data set at the $\alpha = 0.05$ level are given in Table 1. Some promising procedures proposed by Hawkins (1981), Koziol (1982), Paulson et al. (1987), and Singh (1993) were studied by the authors and the results noted in Mecklin (2000). However, these five procedures used empirical critical values. In general, it would be more convenient and practical to be able to use an asymptotic critical value that is accurate even for small values of n and p. For this reason, we limit ourselves here to the 8 procedures given in Table 1.

Srivastava and Hui (1987) noted that an advantage of using asymptotic null distributions is that specialized extensive tables are not needed and the inevitable need to either interpolate critical values or simulate the critical value when the sample size does not exactly match the tabled values is avoided. However, Romeu and Ozturk (1993) and Young et al. (1995) point out that the use of asymptotic null distributions is often quite conservative and results in a loss of power. They would suggest re-determining the critical values empirically via simulation, but we feel that the advantage of using an easily obtained asymptotic critical value that performs well for even small values of p and/or n outweighs the gain in power that can be had by using empirical critical values obtained via simulation. Many potential users of a test of MVN will have neither the necessary skills nor the inclination to empirically determine critical values each time they apply the test.

In a Monte Carlo study, it is important to carefully choose the distributions to be simulated. Many previous Monte Carlo studies have been criticized by Hampel et al. (1985) for being "wasteful and superfluous" and in Horswell (1990) for being "haphazardly selected". Another limitation of some past simulation studies was to only consider multivariate distributions that were merely composed of marginal components independently and identically distributed from some common univariate distribution. However, as Ward (1988) pointed out, uncorrelated variables in multivariate analysis are neither common nor interesting. Thus, any reasonable

Table 2. Multivariate Distributions Simulated

Distribution	Comments
Multivariate Normal	Null hypothesis is true. Symmteric and mesokurtic.
Mixtures of Multivariate Normals	15 different mixtures were considered, with 3 levels of mixing and 5 configurations of means and variances. Mixing level 1 (90%/10%) is mild contamination and is skewed and leptokurtic. Mixing level 2 (78.8675%/21.1325%) is moderate contamination and is skewed and mesokurtic. Mixing level 3 (50%/50%) is severe contamination and is symmetric and platykurtic.
Multivariate t (10 *df*)	Symmetric and mildly leptokurtic.
Multivariate Cauchy	Symmetric and platykurtic.
Pearson Type II	Two variations; one that is a very mild deviation from MVN and another that is more platykurtic.
Multivariate Uniform	Symmteric but highly platykurtic.
Multivariate χ^2 (1 *df*)	Heavily skewed.
Multivariate lognormal	Heavily skewed.
Knitchnine	A distribution with normal marginals but joint non-normality.
Generalized Exponential Power	A non-normal distribution such that the values for Mardia's skewness and kurtosis are equal to the values found for MVN.

study comparing the power of tests of multivariate normality will consider multivariate distributions with correlated components. The multivariate distributions used for this study, given in Table 2, fulfilled this criterion and many were generated with algorithms found in Johnson (1987).

4. RESULTS OF SIMULATION STUDY

The first distribution simulated was the MVN distribution. In this case, the null hypothesis is true, so each test should reject at about the nominal rate of 5%. A rejection rate far above the 5% level would indicate a problem with either the Type I error rate, the accuracy of the asymptotic critical value for certain combinations of n and p, or the accuracy of previously determined empirical critical values. The performance of the 13 tests against the MVN distribution is found in Table 3. Notice that the Mardia-Foster test had a very unreliable performance, with the observed Type I error rate ranging from 0% to 100%. Three other procedures (Mudholkar-Srivastava-Lin, Romeu-Ozturk, and Mardia-Kent) also had a maximum observed rejection rate of over 10%. Mardia's test of multivariate skewness was always slightly liberal, while Mardia's kurtosis and Henze-Zirkler were always slightly conservative. Royston's test had the smallest range of observed Type I error (4.7%-5.3%).

Table 3. Empirical Type I Error Rate Against the Multivariate Normal Distribution

MVN	$n = 25$				$n = 100$			
Test	$p = 2$	$p = 3$	$p = 4$	$p = 5$	$p = 2$	$p = 3$	$p = 4$	$p = 5$
Skew	6.3	6.7	6.6	6.2	5.4	6.1	6.4	6.0
Kurt	0.9	0.5	0.4	1.0	3.1	2.9	3.5	3.5
M-F	16.6	99.9	0.0	0.0	10.8	53.4	97.3	100.0
M-K	4.2	5.0	4.1	3.6	6.2	8.0	9.1	10.2
H-Z	3.8	3.0	2.6	2.2	4.3	4.2	3.7	3.8
Roy	4.9	5.2	4.7	4.8	4.8	5.2	5.3	4.7
MSL	4.9	8.2	12.7	17.0	6.4	14.0	20.0	24.6
R-O	4.9	7.0	9.3	10.6	4.5	7.2	9.4	11.7

With 10000 simulations, the largest possible standard error for the proportion of rejections of MVN is

$$\sigma_{\bar{p}} = \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = \sqrt{\frac{0.5(1-0.5)}{10000}} = 0.005 \qquad (4.1)$$

So an approximate 95% confidence interval for the proportion of rejections is ± 2 standard errors, or $\pm 1\%$. Four procedures (Mardia-Foster, Mardia-Kent, Mudholkar-Srivastava-Lin, Romeu-Ozturk) had occasions where the Type I error rate more than twice exceeded the nominal level of 5%. Thus, with the existence of 4 other procedures with acceptable Type I error rates, these four procedures will neither be considered further nor recommended for use. The maximum standard error of the proportion shows that this criterion for elimination of procedures is actually very conservative.

Five different symmetric distributions from the elliptically contoured family were considered; the multivariate uniform, multivariate t (with 10 *df*), multivariate Cauchy, and two members of the Pearson Type II family. As one would expect, Mardia's skewness had virtually no power against symmetric alternatives. The Henze-Zirkler test and Royston's test were generally powerful against the multivariate uniform, but had minimum power of around 40% when $p = 5$ and $n = 25$, while Mardia's test of kurtosis had power ranging from 89% to 100%. Results for the multivariate uniform distribution are given in Table 4. Results for both Pearson Type II distributions were extremely similar to the uniform. The Pearson Type VII family, of which both the multivariate t and Cauchy are members, represent very mild departures from normality. In general, the rate of rejection was very low for these distributions, generally below 10%.

Both the multivariate lognormal and chi-square distributions are heavily skewed distributions that are drastic departures from normality. Here, the power of the tests of MVN should be very high, close to 100%. For the multivariate chi-square (with 1 *df*), Mardia's skewness, Royston's test, and the Henze-Zirkler procedure all had power of at least 99%, while the power of Mardia's kurtosis dipped as low as 70% when $n = 25$. As sample size increased to 100, the power of all procedures was at least 99.9%.

Empirical power was also very high for the multivariate lognormal distribution. Mardia's kurtosis had the lowest minimum power of 82.5%. Mardia's skewness,

TABLE 4. Empirical Power Against the Multivariate Uniform Distribution

MVN Test	$n = 25$				$n = 100$			
	$p = 2$	$p = 3$	$p = 4$	$p = 5$	$p = 2$	$p = 3$	$p = 4$	$p = 5$
Skew	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Kurt	89.0	95.8	97.2	96.8	100.0	100.0	100.0	100.0
H-Z	100.0	99.8	72.3	39.6	100.0	100.0	100.0	100.0
Roy	100.0	66.5	50.7	40.3	100.0	100.0	100.0	100.0

TABLE 5. Empirical Power Against the Multivariate Lognormal Distribution

MVN Test	$n = 25$				$n = 100$			
	$p = 2$	$p = 3$	$p = 4$	$p = 5$	$p = 2$	$p = 3$	$p = 4$	$p = 5$
Skew	99.8	99.7	99.9	100.0	100.0	100.0	100.0	100.0
Kurt	82.5	90.3	93.7	95.8	100.0	100.0	100.0	100.0
H-Z	99.1	99.7	99.9	100.0	100.0	100.0	100.0	100.0
Roy	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Royston, and Henze-Zirkler all had a minimum power of at least 98%. Table 5 contains results from the lognormal distribution.

Results from the other distributions were generally less interesting and were not as useful in comparing the quality of the different tests of MVN against each other. A total of fifteen different mixtures of multivariate normal distributions were considered, which had various combinations of equal/unequal means, equal/unequal variances, levels of mixing, and amount of correlation between variables. Since all of the normal mixtures considered had little or no separation in the mean vectors, the tests had low power in all situations.

A version of the Knintchine distribution, described in Johnson (1987), was considered. It has the unusual property of marginal distributions that are univariate normal yet the joint distribution is not MVN. Since this distribution is symmetric, Mardia's skewness had virtually zero power. Mardia's kurtosis performed modestly when $n = 25$ (empirical power of 1.8% to 24.8%), at a considerably higher rate when $n = 50$ (60.5% to 87.6%), and almost always rejected MVN when $n = 100$ (99.7% to 100%). The remaining procedures performed similarly.

The generalized exponential power distribution, described in Johnson (1987) is another theoretically fascinating multivariate distribution. The member of this family used in this simulation study had the property of having the same values for multivariate skewness and kurtosis as the MVN distribution. Most tests had quite low power against this distribution, particularly Mardia's kurtosis.

5. DISCUSSION OF RESULTS

Many researchers, such as Bozdogan and Ramirez (1997), Tsai and Koziol (1988), Horswell (1990), Horswell and Looney (1992), Kariya and George (1995), Looney (1995), and Mudholkar et al. (1995) have lamented the widespread neglect given to testing the assumption of normality in multivariate analysis. Any attention given towards assessing the assumptions of a statistical procedure, such as the assumption of MVN, is time well spent.

This study found that no single procedure was the most powerful in every situation. This corroborated previous research by Gnanadesikan (1977), Koziol (1986), Ward (1988), Horswell and Looney (1992), Romeu and Ozturk (1993), Looney (1995), Young et al. (1995), and Bogdan (1999). However, no previous investigation considered all of the procedures that we studied. Of the 8 procedures selected for use in this study, 4 of them (Mardia-Foster, Mardia-Kent, Mudholkar-Srivastava-Lin, Romeu-Ozturk) had an empirical Type I error rate against the multivariate normal distribution that exceeded 10% in some situations. Hence, especially with the availability of other procedures with good Type I error control, these procedures are not recommended.

Of the four remaining procedures, Mardia's skewness was found to be liberal, rejecting the null hypothesis when data were generated from the MVN distribution at a rate slightly higher than 5%. On the other hand, Mardia's kurtosis and Henze-Zirkler were somewhat conservative. Royston's test consistently rejected at nearly the nominal 5% level.

Of the four remaining tests of MVN, situations existed in which some of these procedures had power that was appreciably less than their competitors. Against heavily skewed distributions such as the chi-square or lognormal, Mardia's kurtosis had power as low as 70-80% while Royston's test, the Henze-Zirkler test, and Mardia's skewness have power of virtually 100%. However, Mardia's test of multivariate skewness was a very poor test of MVN, in the sense that the empirical power was virtually nil, against symmetric alternatives such as distributions drawn from the elliptically contoured family.

There were some limitations in this study. First, it would be preferable to determine the power functions analytically, rather than via simulation. However, the true distributions of most of the test statistics and of the associated power functions are mathematically intractable. Thus, the Monte Carlo study was necessary for comparing error rates and power of the MVN tests.

Of the fifty or so tests of MVN in the literature, only eight procedures were considered. There was ample justification for the omission of the vast majority of tests. Some tests rely on empirical critical values, which are not readily attainable for many users. Several tests have been shown in previous studies to be not as powerful as other available tests. Some procedures are either computationally excessive or have algorithmic problems. Certain tests are only available for the bivariate case. Several recently developed tests are designed to be 'locally best' against a specific family of alternative distributions. The goal of this study was to find a test that is generally powerful against a wide range of alternatives. Thus, tests that focus on a limited range of alternatives were excluded from this study. A handful of tests are geared towards particular forms of data (i.e. time series data) and were also excluded.

6. CONCLUSION

Some procedure for assessing the assumption of multivariate normality should be used, even if the subsequent multivariate analyses are robust to violations of MVN. If one is going to rely on one and only one procedure, the Henze-Zirkler test is recommended. This recommendation is based both upon the acceptable (if slighly conservative) Type I error results and power that is either comparable or superior to other procedures against the entire breadth of distributions considered in this study.

Another procedure that had similar power to the Henze-Zirkler and did not suffer from any serious weaknesses was Royston's extension of the familiar Shapiro-Wilk test. The simulation results presented here show Royston's test to have very good Type I error control and power very similar to Henze-Zirkler. However, some theoretical concerns exist for Royston's procedure. Unlike the Henze-Zirkler test, it is not consistent against all alternatives. Further, Royston's test involves a rather ingenious correction for the correlation between the variables in the sample. Srivastava and Hui (1987) felt the correction was not adequately justified, while Romeu and Ozturk (1993) found that Royston's test performed poorly when the variates were highly correlated. Romeu and Ozturk (1993) considered intercorrelations as high as 0.9, while more modest intercorrelations of 0.2 and 0.5 were used in this study.

It was originally suggested by Csörgö (1989) that a strategy for assessing MVN might consist of the use of a procedure that is both consistent and powerful to formally test the null hypothesis of MVN, followed up with less formal procedures if normality was rejected. Based upon the results of this simulation study, the use of the Henze-Zirkler procedure is recommended as the method of choice for assessing goodness-of-fit to a hypothesized MVN distribution. Since the Henze-Zirkler test statistic is not helpful in diagnosing the type of non-normality, a rejection of MVN should be complemented with graphical procedures such as the chi-square plot described in Healy (1968) and the descriptive use of multivariate measures of skewness and kurtosis derived in Mardia (1970). Based upon these supplemental results, the most appropriate next step for the multivariate analysis could be pursued.

References

[1] Andrews, D., R. Gnanadesikan, and J. Warner (1973). *Methods for Assessing Multivariate Normality*, Volume 3 of *Proceedings of the International Symposium on Multivariate Analysis (ed. P.R. Krishnaiah*, pp. 95–116. New York: Academic Press.

[2] Baringhaus, L. and N. Henze (1988). A consistent test for multivariate normality. *Metrika 35*, 339–348.

[3] Bogdan, M. (1999). Data driven smooth tests for bivariate normality. *Journal of Multivariate Analysis 68*, 26–52.

[4] Bozdogan, H. and D. Ramirez (1997). Testing for model fit: Assessing Box-Cox transformations of data to near normality. *Computational Statistics Quarterly 3*, 203–213.

[5] Csörgö, S. (1986). Testing for normality in arbitrary dimension. *Annals of Statistics 14*, 708–723.

[6] Csörgö, S. (1989). Consistency of some tests for multivariate normality. *Metrika 36*, 107–116.

[7] DeCarlo, L. (1997). On the meaning and use of kurtosis. *Psychological Methods 2*, 292–307.

[8] Epps, T. and L. Pulley (1983). A test for normality based on the empirical characteristic function. *Biometrika 70*, 723–726.

[9] Gnanadesikan, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations*. New York: Wiley.

[10] Hampel, F., E. Ronchetti, P. Rousseeuw, and W. Stahel (1985). *Robust Statistics: The Approach Based on Influence Functions*. New York: Wiley.

[11] Hawkins, D. (1981). A new test for multivariate normality and homoscedasticity. *Technometrics 23*, 105–110.

[12] Healy, M. (1968). Multivariate normal plotting. *Appl. Statist. 17*, 157–161.

[13] Henze, N. and B. Zirkler (1990). A class of invariant consistent tests for multivariate normality. *Communs. Statist. Theory Meth. 19*, 3595–3618.

[14] Horswell, R. (1990). *A Monte Carlo comparison of tests of multivariate normality based on multivariate skewness and kurtosis*. Ph. D. thesis, Louisiana State University.

[15] Horswell, R. and S. Looney (1992). A comparison of tests for multivariate normality that are based on measures of multivariate skewness and kurtosis. *Journal of Statistical Computation and Simulation 42*, 21–38.

[16] Johnson, M. (1987). *Multivariate Statistical Simulation*. New York: Wiley.

[17] Johnson, R. and D. Wichern (1992). *Applied Multivariate Statistical Analysis*. Englewood Cliffs, NJ: Prentice Hall.

[18] Kariya, T. and E. George (1995). Locally best invariant tests for multivariate normality. *Sankhya 57*, 440–451.

[19] Kendall, M. and A. Stuart (1977). *The Advanced Theory of Statistics* (4th ed.), Volume 1. Cambridge, UK: Cambridge University Press.

[20] Koziol, J. (1982). A class of invariant procedures for assessing multivariate normality. *Biometrika 69*, 423–427.

[21] Koziol, J. (1983). On assessing multivariate normality. *J. R. Statist. Soc. B 45*, 358–361.

[22] Koziol, J. (1986). Assessing multivariate normality: A compendium. *Communs. Statist. Theory Meth. 15*, 2763–2783.

[23] Koziol, J. (1993). Probability plots for assessing multivariate normality. *The Statistician 42*, 161–174.

[24] Looney, S. (1995). How to use tests for univariate normality to assess multivariate normality. *The American Statistician 39*, 75–79.

[25] Mardia, K. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika 57*, 519–530.

[26] Mardia, K. (1980). *Tests of univariate and multivariate normality*, Volume 1 of *Handbook of Statistics (ed. P.R. Krishnaiah)*, pp. 297–320. Amsterdam: North Holland.

[27] Mardia, K. and K. Foster (1983). Omnibus tests of multinormality based on skewness and kurtosis. *Communications in Statistics 12*, 207–221.

[28] Mardia, K. and J. Kent (1991). Rao score tests for goodness of fit and independence. *Biometrika 78*, 355–363.

[29] Mardia, K., J. Kent, and J. Bibby (1979). *Multivariate Analysis*. New York: Academic Press.

[30] Mecklin, C. (2000). *A Comparison of the Power of Classical and Newer Tests of Multivariate Normality*. Ph. D. thesis, University of Northern Colorado.

[31] Mudholkar, G., D. Srivastava, and C. Lin (1995). Some p-variate adaptations of the Shapiro-Wilk test of normality. *Communications in Statistics: Theory and Methods 24*, 953–985.

[32] Ozturk, A. and E. Dudewicz (1992). A new statistical goodness-of-fit test based on graphical representation. *Biometrical Journal 34*, 403–427.

[33] Paulson, A., P. Roohan, and P. Sullo (1987). Some empirical distribution function tests for multivariate normality. *Journal of Statistical Computation and Simulation 28*, 15–30.

[34] Rencher, A. (1995). *Methods of Multivariate Analysis.* New York: Wiley.

[35] Romeu, J. and A. Ozturk (1993). A comparative study of goodness-of-fit tests for multivariate normality. *Journal of Multivariate Analysis 46*, 309–334.

[36] Romeu, J. and A. Ozturk (1996). A new graphical test for multivariate normality. *American Journal of Mathematical and Management Science 16*, 5–26.

[37] Royston, J. (1983). Some techniques for assessing multivariate normality based on the Shapiro-Wilk w. *Applied Statistics 32*, 121–133.

[38] Shapiro, S. and M. Wilk (1965). An analysis of variance test for normality. *Biometrika 52*, 591–611.

[39] Singh, A. (1993). *Omnibus robust procedures for assessment of multivariate normality and detection of multivariate outliers*, pp. 445–488. Multivariate Environmental Statistics (eds. G.P. Patil and C.R. Rao. Amsterdam: North Holland.

[40] Small, N. (1980). Marginal skewness and kurtosis in testing multivariate normality. *Applied Statistics 29*, 85–87.

[41] Srivastava, M. and T. Hui (1987). On assessing multivariate normality based on Shapiro-Wilk w statistic. *Statistics and Probability Letters 5*, 15–18.

[42] Tsai, K. and J. Koziol (1988). A correlation type procedure for assessing multivariate normality. *Communications in Statistics: Computation and Simulation 17*, 637–651.

[43] Ward, P. (1988). *Goodness-of-fit tests for multivariate normality.* Ph. D. thesis, University of Alabama.

[44] Young, D., S. Seaman, and J. Seaman (1995). A comparison of six test statistics for detecting multivariate nonnormality which utilize the multivariate squared-radii statistic. *Texas Journal of Science 47*, 21–38.