Psychology Science, Volume 48, 2006 (4), p. 419-435

Multinormality and symmetry: a comparison of two statistical tests

ALEXANDER VON EYE[1], MAXINE VON EYE[2] & G. ANNE BOGAT[1]

Abstract

Multinormal distributions are symmetric. The degree of deviations from axial symmetry can be assessed using the well known Bowker test. A recently proposed test (von Eye & Bogat, 2004; von Eye & Gardiner, 2004) is based on comparing the observed frequencies in sectors of the multivariate space with the corresponding expected frequencies that were estimated based on multinormality. Because this test is an omnibus test of multinormality, it should also be sensitive to deviations from axial symmetry. In this article, we describe the results of simulations that were performed on four types of bivariate distributions: normal, uniform, inverse Laplace-transformed, and cube-root transformed. As expected, the Bowker test showed that inverse Laplace-transformed distributions are likely to show deviations from axial symmetry. None of the other distributions was asymmetric. The new omnibus test of multinormality exhibited 100 % sensitivity to violations of axial symmetry, but was also sensitive to elevated skewness and kurtosis. Thus, it also flagged the uniform and the cube root-transformed distributions as deviating from multinormality. Results also show that the Bowker test is sensitive only to violations of axial symmetry.

Key words: multinormality; testing multinormality; symmetry; Monte Carlo

---

[1] Address correspondence concerning this article to Alexander von Eye, Michigan State University, 316 Psychology Building, East Lansing, MI 48824-1116, USA; e-mail: voneye@msu.edu.

[2] University of Cambridge, UK

Multinormality is a condition required for proper application of parametric multivariate statistical procedures. There exists a large number of tests of multinormality including, e.g., Mardia's (1970, 1980) tests of multivariate skewness and kurtosis. Recently, two tests of multinormality were proposed that are based on segmenting the variables under study such that the segments are either equidistant on the variables or equiprobable (von Eye and Bogat, 2004; von Eye & Gardiner, 2004). The first of these tests examines each of the sectors that result from crossing the segmented variables, asking whether the number of cases found for a sector corresponds to the number expected under the assumption of a multinormal distribution. The second test is based on a summary deviance score for all sectors. It thus examines the entire distribution. It is obvious that sectors mirrored at the centroid will be expected to contain the same number of cases. In other words, multinormality implies multivariate point symmetry with respect to the centroid. Accordingly, sectors mirrored at the main diagonal will be expected to contain the same number of cases, because multivariate normality also implies axial symmetry.

In this study, we ask whether the new tests, in particular the second, which, recently, was shown to be omnibus to violations of skewness, kurtosis, and a series of distributional transformations (von Eye, 2005), can also be used to examine multivariate axial symmetry. We present a simulation study in which data are studied that differ in the degree to which they deviate from a multinormal distribution, and in the type of violation. In particular, the simulated data differed in axial symmetry. It will be asked whether the second of the new tests is also sensitive to symmetry characteristics. To answer the question, the new omnibus test is compared to Bowker's test (1948).

1. The new tests of multinormality

The tests proposed by von Eye and Bogat (2004) and von Eye and Gardiner (2004) can be seen as multivariate extensions of the univariate X^2-test of normality that is described in many introductory statistics text books (e.g., Glass & Hopkins, 1984, Ch. 14.7). For these tests, the range of the variable under study is partitioned into two or more segments. The probability of each segment is determined under the assumption of a univariate normal distribution. For the multivariate case, the analogous procedure can be described algorithmically as follows[3].

(1) *Split each of the d variables under study into 2 or more segments*. For the jth variable, we obtain $c_j > 1$ segments, with j = 1, ..., d.

(2) *Cross the segmented variables*. Crossing all segmented variables yields a *d*-dimensional cross-classification with $\prod_{j=1}^{d} c_j$ sectors.

(3) *Calculate the probability of each sector*. Consider the univariate case first. Let the boundaries of a segment be z_k and z_{k+1}. The area under the normal curve for this segment is

---

[3] The following paragraphs borrow heavily from von Eye & Bogat (2004).

$$p(z_{k+1}) - p(z_k) = \int_{-\infty}^{z_{k+1}} \Psi(z)dz - \int_{-\infty}^{zk} \Psi(z)dz\,,$$

where k indexes the lower end and $k + 1$ the upper end of the segment. In the multivariate case, cases sit in the sectors that were created by crossing all segmented variables. Each of these sectors has boundaries z_i^1 and z_{i+1}^1 on the first variable, z_j^2 and z_{j+1}^2 on the second variable, ..., and z_l^d and z_{l+1}^d on the dth variable, where the subscripts indicate the segments and the superscripts indicate the variables. The probability of being located in the sector with these boundaries is

$$p(z_i^1 - z_{i+1}^2, z_j^2 - z_{j+1}^2, \ldots z_l^d - z_{l-1}^d) = \int_{z_i^1}^{z_{i+1}^1} \int_{z_j^2}^{z_{j+1}^2} \ldots \int_{z_l^d}^{z_{l+1}^d} \Psi(z^1, z^2, \ldots, z^d) \quad dz^1\, dz^2 \ldots dz^d\,.$$

Genz (1992) presented a computational solution for this equation (cf. Gupta, 1963; for solutions for bivariate analyses see Maydeu & Olivares, 2001; Seidler & Formann, 1980; the software packages S plus and Mathematica can be used to solve this equation; a Fortran subroutine is available from Genz, 1992). The sectors defined this way are multivariate-rectangular. Somerville (1998) proposed a solution for convex, that is, ellipsoid sectors. For the following simulations, we use Genz's solution. We label the individual sector $s_{i,\,j,\ldots,l}$, and abbreviate the probability of sitting in this d-dimensional right-angled sector with $p_{i,\,j,\ldots,l}$.

(4) *Estimate expected sector frequencies*. The expected frequency of objects in Sector $s_{i,j,\ldots,l}$ is $e_{i,j,\ldots,l} = Np_{i,j,\ldots,l}$. The next step performs the first test proposed by von Eye and Bogat (2004) and von Eye and Gardiner (2004). It examines each individual sector.

(5) *Perform sector-specific tests*. To identify locations of violations of multinormality, for each Sector, $s_{i,j,\ldots,l}$, the observed frequency of objects, $o_{i,j,\ldots,l}$, is compared with the corresponding expected frequency, $e_{i,j,\ldots,l}$, under the null hypothesis $E[o_{i,j,\ldots,l}] = e_{i,j,\ldots,l}$. If this comparison suggests that a sector contains significantly more or fewer objects than expected based on the assumption of a multinormal distribution of the d variables under study, this sector displays a violation of multivariate normality. Therefore, the assumption of multivariate normality is rejected at least for this sector. Many tests exist that are suitable for the present purpose (von Eye, 2002). Von Eye and Bogat (2004) proposed using the well-known Pearson X^2-component test, for three reasons. First, the X^2-components, also termed *standardized residuals*, are known to have desirable properties for the analysis of individual cells of a multivariate cross-classification (von Eye, 2002; von Weber, von Eye, & Lautsch, 2004). Second, the component test is a straightforward extension of the univariate χ^2-test. Third, the components sum up to an omnibus test statistic. The test statistic for the individual sector is

$$X^2_{i,j,\ldots,l} = \frac{(o_{i,j,\ldots,l} - e_{i,j,\ldots,l})^2}{e_{i,j,\ldots,l}}$$

with $df = 1$. Because of the possibly large number of tests, it is advisable to protect the significance threshold α. Most popular is the Bonferroni procedure which takes only the total number of tests into account. The Bonferroni-protected threshold is $\alpha^* = \alpha / \prod_{j=1}^{d} c_j$. More efficient procedures have been proposed by Holm (1979) or Keselman, Cribbie, and Holland (1999). These procedures yield protected α-thresholds that are less prohibitive.

The use of standardized residuals to test sector-specific null hypotheses can be based on the following asymptotic behavior of the residuals. Let the number of cells in the table go to infinity as the number of segments increases, let $p_{ij} = p_{i.}\, p_{.j}$ (variable independence), and $N \to \infty$. Under these conditions, $p_{ij} = p_{i.}\, p_{.j} \to 0$. Therefore, $(1 - \frac{1}{\hat{m}_{ij}}) \to 1$, $e_{ij} = \frac{n_{ij} - \hat{m}_{ij}}{\sqrt{\hat{m}_{ij}(1 - \frac{1}{\hat{m}_{ij}})}} \to \frac{n_{ij} - \hat{m}_{ij}}{\sqrt{\hat{m}_{ij}}}$, and $\sqrt{X^2}$ is normally distributed. Only if the model fits will the variance be less than 1.

(6) *Omnibus test.* The following is the second test proposed by von Eye and Bogat (2004) and von Eye and Gardiner (2004). The sum of the X^2-components yields the test statistic

$$X^2 = \sum_{i,j,\ldots,l} \frac{(o_{i,j,\ldots,l} - e_{i,j,\ldots,l})^2}{e_{i,j,\ldots,l}} = \sum_{i,j,\ldots,l} X^2_{i,j,\ldots,l} .$$

This statistic can be used as a test of whether, overall, the cross-classification of segments follows a multinormal distribution. The test has $df = \left(\prod_{j=1}^{d} c_j\right) - 2d - d_{\text{cov}} - 1$, where c_j is the number of segments of the jth variable (for more detail for the univariate case, see Dahiya & Gurland, 1972, 1973). The term d_{cov} indicates the number of correlations (or covariances) taken into account. Typically, $d_{\text{cov}} = \binom{d}{2}$; that is, all covariances are taken into account. For the following simulations, we focus on the omnibus test, and compare it to Bowker's (1948) X^2-test of axial symmetry.

2. Bowker's Test of Axial Symmetry

Bowker's X^2-test allows one to examine square contingency tables with respect to axial symmetry. If, for the cells of an I x J table, with $I = J$, $o_{ij} = o_{ji}$, for $i \neq j$, the table is said to possess *axial symmetry*. Bowker's X^2-test is

$$X^2 = \sum_i \sum_j \frac{(o_{ij} - o_{ji})^2}{o_{ij} + o_{ji}}$$

for $i > j$. The test is distributed approximately as χ^2 with $\binom{I}{2} = \frac{I(I-1)}{2}$ degrees of freedom.

This test has also been discussed in the context of nonstandard log-linear models (von Eye & Spiel, 1996). Multivariate versions of Bowker's test are straightforward. In the present context, we illustrate the characteristics of this test using I x J tables.

3. The simulation

The simulation was performed to examine the behavior of Bowker's test and the new omnibus test under conditions of various variable characteristics. All simulation runs were performed using FORTRAN 90 programs that were written for this purpose. Specifically, bivariate distributions were created that exhibited the following four distributional characteristics (cf. Fleishman, 1978; Vale & Maurelli, 1983).

(1) *Normal distribution*. To create univariate normally distributed data, the generator GASDEV from the Numerical Recipes FORTRAN collection (Press, Flannery, Teukolsky, & Vetterling, 1989) was used. This generator returns a normally distributed deviate with zero mean and unit variance (see also Sicking, 1994). It is based on the function RAN1, also provided in the recipe collection. RAN1 returns Gaussian pseudo random deviates. The generator requires a user-specified seed. This seed was created using the library MSFLIB that is available in the MS Fortran Power Station. The distributions created this way are expected to be symmetric.

(2) *Uniform distribution*. To create uniformly distributed data, the generator RANDOM, available in the Power Station's PortLib function pool, was used. This generator returns pseudo random numbers, z, from the interval $0 \leq z < 1$. The algorithm used is a prime modulus M multiplicative congruential generator (Park & Miller, 1988). The distributions created this way are expected to be symmetric.

(3) "*Inverse Laplace-transformed.*" The Laplace probability distribution is

$$f(x) = \frac{1}{2\beta} \exp\left[-\left(\frac{|x - \alpha|}{\beta} \right) \right]$$

for $x < \alpha$, $-\infty < x < \infty$, and $\beta > 0$. This distribution has a mean of zero, a skewness of zero, and an elevated kurtosis. For $\beta = 1$ and x centered scores, the probability distribution becomes

$$y = \frac{\exp^{-|x|}}{2}$$

The Laplace distribution is a one-peaked, symmetric distribution, as illustrated in Figure 1.

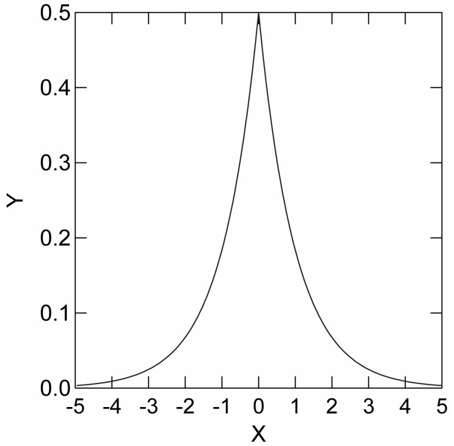

Figure 1:
Illustration of Laplace distribution

A uniform distribution has no skew but an increased kurtosis. In contrast, a Laplace distribution is known to be *supergaussian*, that is, leptokurtic. The Laplace transformation is linear. It is used in, for instance, Process Dynamics to convert time-domain relationships to a set of equations expressed in terms of the Laplace operator (Tham, 1999). One application in data analysis aims at changing the kurtosis of a distribution. Performing an inverse Laplace transformation on a uniform distribution should, therefore, result in a distribution with elevated kurtosis and possibly elevated skewness. However, the Laplace function has no straightforward inverse.

Therefore, we propose and perform the following transformation. First, a uniform distribution of scores *y* was created using the RANDOM generator described above. Second, the values were transformed to be within the interval {0, 0.5}. Third, a logarithmic transformation was performed, that is, $x = \ln |2y|$. Finally, a random half of the resulting scores was set to $-x$. The scatterplot matrix in Figure 2 illustrates this procedure on a sample of 100 random data (for this example, SYSTAT's URN generator was used, which returns uniformly distributed random numbers within the interval {0, 1}).

Figure 2 shows that the uniform distribution created by URN is reasonably close to rectangular. It has a slightly elevated kurtosis (see Table 1, below), but its skewness does not differ from zero. Inserting into the Laplace function results in the distribution in the second row of the scatterplot matrix. This distribution also has a slightly elevated kurtosis, and its skewness is not statistically different than zero. The substitute for the inverse Laplace distribution results in the distribution shown in the third row of scatterplots in Figure 2. As Table 1 documents, this distribution has a slightly elevated skewness and an elevated kurtosis. Please note that this transformation results in a kurtosis with a positive sign. The kurtosis of the uniform distribution was negative. In other words, this transformation changed the distribution from being heavy-tailed to heavy around the belt line. Because of their elevated skewness, we expect the distributions that result from this transformation to deviate from axial symmetry.

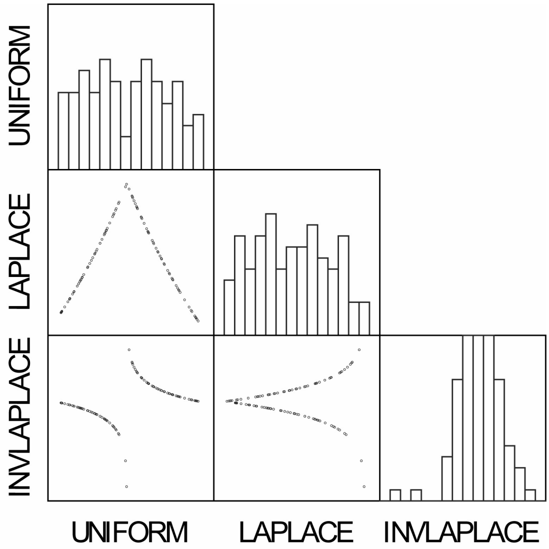

Figure 2:
Illustration of transformation to create a substitute for the inverse Laplace distribution

Table 1:
Descriptive statistics of the variables plotted in Figure 2

	UNIFORM	LAPLACE	INVLAPLACE
Number of cases	100	100	100
Minimum	-0.456	0.305	-5.716
Maximum	0.494	0.499	3.471
Mean	-0.004	0.396	0.010
Standard Deviation	0.279	0.054	1.344
Skewness (G1)	0.067	0.041	-0.663
SE Skewness	0.241	0.241	0.241
Kurtosis (G2)	-1.156	-1.108	3.027
SE Kurtosis	0.478	0.478	0.478

Changes in kurtosis can also be achieved using other, simpler transformations than the Laplace and the inverse Laplace transformations. Examples include the Fourier transformation which uses trigonometric functions. For the present simulations, the inverse Laplace transformation was used (1) because of the importance of this transformation in system dynamics, and (2) to illustrate the effectiveness of the variant of the inverse transformation procedure proposed in this article.

(4) *Cube root transformation*. This transformation was used to create $y = x^{1/3}$ from the uniform x scores. Considering that the scores that were cube root-transformed had no skewness and an only slightly elevated kurtosis, the resulting scores should have neither. Indeed the sample data from Figure 2 and Table 1 have a skewness of -0.458 (se = 0.337), and a kurtosis of -0.652 (se = 0.662). Neither value is large or significant. We do not expect the distributions created by this transformation to deviate from axial symmetry.

For the simulation, the following data sets were thus created: (1) normally distributed variates; (2) uniformly distributed variates, ranging from 0 to 1; (3) uniformly distributed variates that were subjected to the substitute of the inverse Laplace transformation; and (4) uniformly distributed variates that were subjected to the cube root transformation.

In addition to *type of distribution*, the following data characteristics were varied in the simulation:

1. The *sample size* varied from 50 to 800, in steps of 50. Thus, 16 different sample sizes were used.
2. The *number of segments* of each variable varied from 2 to 5, in steps of 1. Thus, 4 different numbers of segments were used. Because of the definition of axial symmetry that underlies the Bowker test, the number of segments was always the same for both variables.

It should be noted that, for the present simulations, the segmenting was performed on the standardized *z*-scale. Using different scaling could have the effect that the symmetry axis of the normal distribution fails to correspond with the main axis of the contingency table studied here, if the variables are correlated.

The resulting design was thus a 4 (type of distribution) x 16 (sample size) x 4 (number of segments) design with 256 cells. The runs were repeated with 5 different seeds. Thus, a data body with results from 1280 runs was available for analysis.

There were two dependent measures. The first indicated whether a particular bivariate distribution was identified by the Bowker test as violating the null hypothesis of axial symmetry. This variable was dichotomous. The second dependent measure was also dichotomous. It indicated whether the new omnibus test found that a bivariate distribution violated symmetry or other data characteristics that the test is sensitive to, e.g., kurtosis-related deviations from a bivariate normal distribution. The same sectors were used for both dependent measures.

In the following sections we ask (1) whether the Bowker test of symmetry detected any deviations from symmetry in the two-dimensional arrays; (2) whether the new omnibus test was also able to detect the asymmetric distributions; (3) whether the detection of asymmetry depends on sample size; (4) whether the detection of asymmetry depends on the number of segments a data set is partitioned in; and (5) whether the detection of asymmetry depends on distribution type. In addition, we ask questions concerning the sensitivity and specificity of the new omnibus test, in comparison with Bowker's test.

4. Results

To answer the questions concerning the effects of the design variables of the simulations, a MANOVA was performed, followed up by univariate *post hoc* tests. The MANOVA contained the fixed-effect factors *Number of Segments*, and *Distribution Type*, and the covariate *Sample Size*. The frequencies with which Bowker's test and the new omnibus test indicated significant X^2 values were used as dependent variables. In the following paragraphs, we first report results of the overall tests. Specifically, we report the results of MANOVA significance tests and the portions of explained variance, based on Wilks' Lambda. Details concerning the individual effects follow in the context of *post hoc* comparisons.

Effects of the Factors of the Simulation. The F-values for the univariate effects of the sample size are given in Table 2. The table shows that the sample size had a significant effect on the number of instances the two tests signaled significant deviations. 3.8% of the variance was explained by the effect of *N*.

The F-values for the univariate effects of the number of segments are given in Table 3. The table shows that the number of segments had a significant effect on the number of instances with which significant deviations were signaled only on the Bowker test. The overall X^2 was not significantly affected. 8.5% of the variance was explained by the effect of the number of segments.

The F-values for the effect of *Distribution Type* are given in Table 4. The table shows that Distribution Type had the strongest effect on the number of instances the two tests signaled deviations. 95.5% of the variance was explained by this effect.

Table 2:
Univariate F-tests for Sample Size, *N*

Source	SS	*df*	MS	F	*p*
SIGPEARS	0.457	1	0.457	41.292	0.000
Error	13.980	1263	0.011		
BOWKER	0.318	1	0.318	7.627	0.006
Error	52.657	1263	0.042		

Table 3:
Univariate F-tests for Number of Segments

Source	SS	*df*	MS	F	*p*
SIGPEARS	0.084	3	0.028	2.517	0.057
Error	13.980	1263	0.011		
BOWKER	4.569	3	1.523	36.528	0.000
Error	52.657	1263	0.042		

Table 4:
Univariate F-tests for Distribution Type

Source	SS	*df*	MS	F	*p*
SIGPEARS	232.615	3	77.538	7004.851	0.000
Error	13.980	1263	0.011		
BOWKER	14.106	3	4.702	112.781	0.000
Error	52.657	1263	0.042		

Table 5:
F-tests for the Distribution Type x Number of Segments Interaction

Source	SS	*df*	MS	F	*p*
SIGPEARS	0.188	9	0.021	1.890	0.050
Error	13.980	1263	0.011		
BOWKER	20.538	9	2.282	54.733	0.000
Error	52.657	1263	0.042		

Table 5 shows the F-values for the *Distribution Type* x *Number of Segments* interaction. The table shows that an interaction between Distribution Type and Number of Segments can be established only for the Bowker test. 28.89% of the variance was explained by this interaction.

The follow-up tests (all Games-Howell tests) to be reported now are based on the univariate ANOVAs that were performed separately for the two dependent measures. For the Bowker test, *N* explained 4.8% of the variance. The effect was significant but small. Indeed, none of the follow-up tests indicated a significant difference. We thus conclude that the sample size did not have a large effect on the Bowker test.

The Bowker test did respond to the number of segments. 6% of the variance was explained by this factor. With the exception of the comparison of 3 with 4 segments, the portion of significant results decreased as the number of segments increased. The reason for this result is the decreasing power that results from increasing the number of segments. The increases were crossed with sample size. Therefore, for each sample size, an increase in the number of segments caused a decrease in power.

The Bowker test did respond strongly to type of distribution. 16.4% of the variance was explained by this factor. The post hoc tests suggest that the third distribution, that is, the one that was substituted to create an inverse Laplace distribution, differed from the other three by causing a strongly increased portion of significant results. The other three distributions had effects that did not differ from each other. This is illustrated in Figure 3. Figure 3 also shows that the base rate for significant results from the Bowker test is below 0.05 when deviations from symmetry are only random, as is the case for the normal, the uniform, and the cube-transformed distributions. The Bowker test can thus be considered a conservative test.

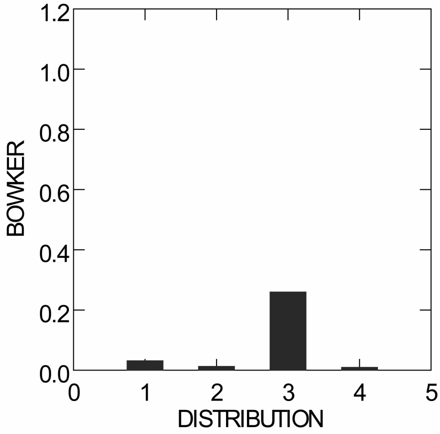

Figure 3:
Portion of significant Bowker tests, by type of distribution, with 1 = normal, 2 = uniform, 3 = inverse Laplace, and 4 = cube root

As far as the new omnibus test is concerned, the effect of sample size explained 15.6% of the variance. This effect is significant. However, as for the Bowker test, none of the *post hoc* tests indicated a significant difference. The number of segments explained only 0.7% of the variance. Here again, although the overall test was significant, none of the *post hoc* tests suggested an individually significant difference. Thus, we conclude that the changes in power that were observed for Bowker's test did not surface for the new omnibus test.

Finally, type of distribution explained 95% of the variance. For both tests under study, this variable had the strongest effect. However, the effect is not the same for the two tests. In contrast to Bowker's test, the new omnibus test is sensitive to the data structures represented by the second, third, and fourth distribution types, that is, the uniform distribution, the inverse Laplace, and the cube root-transformed distributions. In addition, it seems to be equally sensitive to violations that come with these distribution types. This is illustrated in Figure 4. Figure 4 shows that close to 100% of the data situations created by Distributions 2, 3, and 4 resulted in the new omnibus test signaling significant deviations. For the first distribution, that is, for the normal distribution, practically no situation was singled out as significantly deviating. We thus conclude that (1) the normal distribution possesses none of the data characteristics to which the new omnibus test is sensitive, and (2) the test is conservative.

Figure 5 illustrates the interaction effect on the results of the Bowker test. The figure shows the interesting result that the Bowker test is largely unaffected by the number of segments. However, for the inverse Laplace-transformed distributions and even numbers of segments, the number of flagged distributions increases. The reason for this observation is that this distribution also created slightly skewed distributions. As was mentioned above, this had the effects that the table showed a correlation, and the symmetry axis of the normal distribution is not captured by a middle category. This effect seems to be strongest in cases with only 2 x 2 segments. In larger tables, the effect decreases.

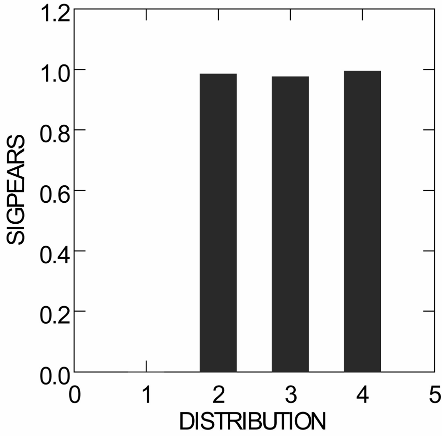

Figure 4:
Portion of significant results for the new omnibus test, by type of distribution, with 1 = normal, 2 = uniform, 3 = inverse Laplace, and 4 = cube root

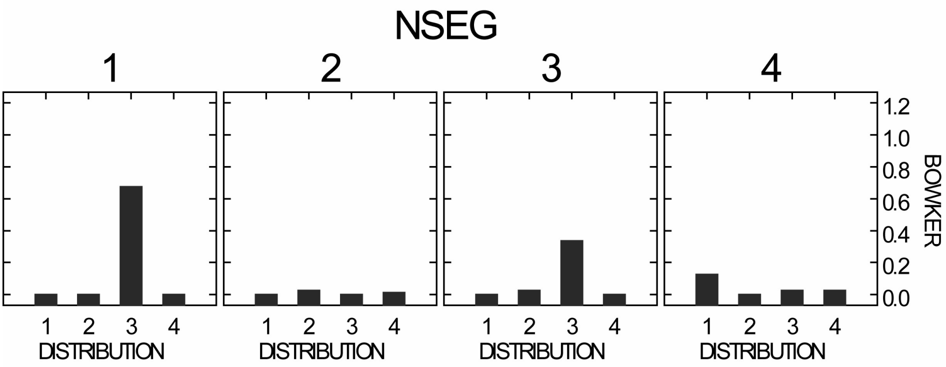

Figure 5:
Portion of significant results for the Bowker test, by type of distribution, with 1 = normal, 2 = uniform, 3 = inverse Laplace, and 4 = cube root, and Number of Segments (note that the label above the graphs number the categories; Number of Segments = Number of Categories + 1)

Specificity and Sensitivity. Based on these results, we now ask questions concerning specificity and sensitivity of the new omnibus test. *Sensitivity* is defined as the portion of cases that do show a characteristic and are correctly classified as showing it. In the present context, we ask, how big is the portion of cases which, based on Bowker's test, do show asymmetry and are correctly classified by the new omnibus test as showing asymmetry. Accordingly, *specificity* is defined as the portion of cases who do not show a characteristic and are correctly classified as not showing it. This can be seen as the *conditional probability* of a negative test result, given that the characteristic under study is absent.

Both, specificity and sensitivity will be discussed in reference to Bowker's test. This is done for two reasons. First, Bowker's test is known to be sensitive to violations of axial symmetry. It thus serves as the benchmark with which the new omnibus test is compared. Second, based on the way the data were created for the present simulations, it was not clear by design which individual data set would violate the null hypothesis of axial symmetry. Therefore again, Bowker's test serves as the reference.

In a first step towards examining the new omnibus test's sensitivity, we cross the two dichotomous dependent measures. The resulting cross-classification appears in Table 6.

Cohen's κ for the data in Table 5 is 0.036. This value is significant. However, it also indicates that general knowledge about the statistical decisions suggested by Bowker's is only 3.6 % better than no knowledge. We therefore ask more specific questions. In particular, we ask whether those bivariate distributions that were identified by the Bowker test as most likely to violate the null hypothesis of axial symmetry are also identified as violating by the new omnibus test. To answer this question, we focus on the distributions generated for the third distribution type, that is, the distributions that result from the substitute for the inverse Laplace distribution. Table 7 displays the resulting cross-classification.

Cohen's κ for the data in Table 6 is even smaller than for the data in Table 5; it is κ = 0.018. This value is also significant. For the current questions, most interesting is the frequency distribution in the second row. This row shows the frequency distribution that is found for the new omnibus test for those cases in which Bowker's test suggests that violations of the null hypothesis of axial symmetry exist. The distribution shows that 100% of the bivariate distributions that were flagged by Bowker's test were also flagged by the new

Table 6:
Cross-classification of the results from the Bowker and the new omnibus tests, aggregated across all distribution types

	new omnibus test		
Bowker test	0	1	Total
0	325	855	1180
1	10	90	100
Total	335	945	1280

Table 7:
Cross-classification of the results from the Bowker and the new omnibus tests, for the inverse Laplace distribution

	new omnibus test		
Bowker test	0	1	Total
0	8	229	237
1	0	83	83
Total	8	312	320

omnibus test. With Bowker's test as a reference, we can therefore state that *the new omnibus test shows 100% sensitivity*.

In contrast, the specificity of the new omnibus test is only 3.38%. Almost 97% of the distributions that Bowker's test had identified as not violating the null hypothesis of axial symmetry were flagged by the new omnibus test. How can this be explained? We discuss three reasons for this discrepancy.

In earlier studies (von Eye, 2005; von Eye & Bogat, 2004; von Eye & Gardiner, 2004), it was shown that the new omnibus test is sensitive to a variety of violations of multinormality. Such violations can exist even if a distribution is symmetric. Imagine, for example, a distribution with strong positive kurtosis, that is, a distribution with more cases close to the mean than expected based on the multinormality assumption. This distribution may well be symmetric, but it violates the multinormality assumption. None of the cases in which this pattern occurs is considered suspicious by Bowker's test. However, these cases will be flagged by the new omnibus test. This applies accordingly to distributions with strong negative kurtosis and other violations of multinormality that conserve axial symmetry. Obviously, the situations created by the current simulation resulted in many more violations of multinormality that are unrelated to axial symmetry than violations that are related to axial symmetry.

A second reason for these differences may be grounded in differential power. This reason, however, is most likely of little importance. Considering that neither test identified even 5 % of the normally distributed data sets as indicating violations, we assume that differences in power are minimal. Both tests are Pearson X^2-tests.

A third reason lies in the test-specific ways of estimating expected cell frequencies. In the two-dimensional environment simulated here, the probability of being located in the sector with boundaries z_i^1 and z_{i+1}^1 on the first variable, z_j^2 and z_{j+1}^2 on the second variable is

$$p(z_i^1 - z_{i+1}^1, z_j^2 - z_{j+1}^2) = \int_{z_i^1}^{z_{i+1}^1} \int_{z_j^2}^{z_{j+1}^2} \Psi(z^1, z^2)\, dz^1\, dz^2$$

for $i \neq j$. The probability of being located in the corresponding sector, that is the sector mirrored at the main diagonal, is

$$p(z_j^1 - z_{j+1}^1, z_i^2 - z_{i+1}^2) = \int_{z_j^1}^{z_{j+1}^1} \int_{z_i^2}^{z_{i+1}^2} \Psi(z^1, z^2)\, dz^1\, dz^2$$

These two probabilities are the same when the bivariate distribution is symmetric, as was illustrated by the uniform and the normal distribution data in the present simulations. In the simulations, variables were independent. In the case of dependent variables, these two probabilities are still the same if the correlations are taken into account for the estimation of expected sector frequencies (see Genz, 1992). In all these cases, Bowker's test and the new omnibus test will indicate that no violations exist.

When kurtosis is extreme and no other violation of normality exists, the new omnibus test will respond, but not Bowker's symmetry test. The reason is that, to perform the symmetry test, the expected frequencies are estimated from the marginals. Thus, heavy or light tails are taken into account. In contrast, the expected frequencies for the new omnibus test are based on the assumption of a multinormal distribution. Therefore, (main) effects such as those that result from extreme kurtosis cause the test to flag a distribution as significantly deviating from normality. Accordingly, root transformations that do not affect skewness will not be flagged by Bowker's test but by the new omnibus test, if they affect kurtosis.

Finally, we ask whether Bowker's test is able to identify violations from multinormality that are unrelated to axial symmetry. To answer this question, we create the same cross-classification as in Table 6, but for the second and the fourth distributions. These are the uniform distribution and the distribution that resulted from calculating the cube root. Table 8 displays the resulting cross-classification for the uniform distribution, and Table 9 displays the cross-classification for the cube root-transformed distribution.

Tables 8 and 9 show corresponding figures. The new omnibus test is highly sensitive to the violations of multinormality created by the uniform and the cube root-transformed distributions. These seem to be mostly violations that lead to significant skewness and kurtosis. Only a very small number of these violations seems to also come with violations of axial symmetry. Therefore, we can also say that Bowker's test is highly specific. No-violations to axial symmetry are unlikely to be flagged, that is, false positives are unlikely.

Table 8:
Cross-classification of the results from the Bowker and the new omnibus tests, for the uniform distribution

	new omnibus test		
Bowker test	0	1	Total
0	5	311	316
1	0	4	4
Total	5	315	320

Table 9:
Cross-classification of the results from the Bowker and the new omnibus tests, for the cube root-transformed distribution

	new omnibus test		
Bowker test	0	1	Total
0	2	315	317
1	0	3	3
Total	2	318	320

5. Discussion

The present article presents the comparison of an omnibus test of deviations from multinormality with a specific test of axial symmetry. The omnibus test proposed by von Eye and Bogat (2004) and von Eye and Gardiner (2004) is unique in the sense that it is tractable for more than two dimensions (cf. Einmahl & McKeague, 2002). In addition, and this is one of the results of the present study, it is not only sensitive to violations of multinormality that materialize in elevated levels of skewness and kurtosis, but also to violations of symmetry as was exemplified using axial symmetry. Other recent tests of multinormality have been shown to be omnibus to deviations from normality that result in skewness or kurtosis (Zhang, 1999). The present results show that the new omnibus test is, in addition, sensitive to 100% of symmetry violations of normality, when compared to Bowker's test.

The present results suggest clearly that Bowker's test is sensitive only to symmetry violations. Bowker's test is marginal-dependent (Goodman, 1991; von Eye & Mun, 2003). The new omnibus test is marginal-free.

We now ask, which of the tests is a researcher to apply? If researchers are interested in violations of multinormality in general, that is, violations of any sort, the new test is probably a good choice. It indicates that violations are present, and researchers then can make decisions concerning the choice of methods of analysis, or resampling (von Eye & Bogat, 2004). The violations thus detected will include violations due to lack of axial symmetry, if they are present.

If, however, researchers are interested exclusively in violations of axial symmetry, Bowker's test is a good choice. This test is sensitive to symmetry violations only. However, considering that the present simulations suggest that the new omnibus test will detect about 100% of the violations that Bowker's test detects, the new test may be an equally good choice. Researchers should keep in mind that the two tests are sensitive to different kinds of symmetry violations. Bowker's test does take main effects into account that may violate the assumption of multinormality. It is thus sensitive to symmetry violations regardless of whether they also imply violations of the multinormality assumption. The new test is sensitive to symmetry violations only if they also violate multinormality assumptions.

References

Bowker, A.H. (1948). A test for symmetry in contingency tables. Journal of the American Statistical Association, 43, 572 - 574.

Einmahl, J.H.J., & McKeague, J.W. (2002). Empirical likelihood based hypothesis testing. stat.fsu.edu/~mckeague/ps/EMKtests-revised-final.pdf

Dahiya, R.C., & Gurland, J. (1972). Person chi-squared test of fit with random intervals. Biometrika, 59, 147 - 153.

Dahiya, R.C., & Gurland, J. (1973). How many classes in the Pearson chi-square test? Journal of the American Statistical Association, 68, 707 - 712.

Fleishman, A. (1978). A method for simulating nonnormal distributions. Psychometrika, 43, 521 - 532.

Genz, A. (1992). Numerical computation of multivariate normal probabilities. Journal of Computational and Graphical Statistics, 1, 141 - 149.

Glass, G.V., & Hopkins, K.D. (1984). Statistical methods in education and psychology, 2nd ed. Englewood Cliffs, NJ: Prentice Hall.

Goodman, L.A. (1991). Measures, models, and graphical displays in the analysis of cross-classified data. Journal of the American Statistical Association, 86, 1085 - 1111.

Gupta, S.S. (1963). Probability integrals of multivariate normal and multivariate t. The Annals of Mathematical Statistics, 34, 792 - 828.

Holm, S. (1979). A simple sequentially rejective multiple test procedure based on a modified Bonferroni test. Scandinavian Journal of Statistics, 6, 65 - 70.

Keselman, H.J., Cribbie, R., & Holland, B. (1999). The pairwise multiple comparison multiplicity problem: an alternative approach to familywise and comparisonwise Type I error control. Psychological Methods, 4, 58 - 69.

Mardia, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. Biometrika, 57, 519 - 530.

Mardia, K.V. (1980). Tests of univariate and mutivariate normality. In P.R. Krishnaiah (Ed.), Handbook of statistics (Vol. 1; pp 279 - 320). Amsterdam: North Holland.

Maydeu-Olivares, A. (2001). Testing categorized bivariate normality with two-stage polychoric correlations. Dept. of Psychology. University of Barcelona. http://www.ub.es/personal/amaydeu/testing%20categorized%20bivariate%20normality.pdf

Park, S.K., & Miller, K.W. (1988). Random number generators: good ones are hard to find. CACM , 31, 1192 - 1201.

Press, W.H., Flannery, B.P., Teukolsky, S.A., & Vetterling, W.T. (1989). Numerical recipes. The art of scientific computing (FORTRAN version). Cambridge: Cambridge University Press.

Seidler, H., & Formann, A.K. (1980). Einige Bemerkungen zur statistischen Auswertung anthropometrischer Daten [Some comments on the statistical analysis of anthropometric data]. Anthropologischer Anzeiger, 38, 270 - 285.

Sicking, F.J. (1994). GASDEV. http://www.astro.rug.nl/~gipsy/sub/gasdev.dc2

Somerville, P.N. (1998). Numerical computation of multivariate normal and multivariate-t probabilities over convex regions. Journal of Computational and Graphical Statistics, 7, 529 - 544.

Tham, M.T. (1999). Process dynamics: Laplace transforms. Lorien.ncl.ac.uk/ming/dynamics/laplace.pdf

Vale, C.D., & Maurelli, V.A. (1983). Simulating multivariate nonnormal distributions. Psychometrika, 48, 465 - 471.

von Eye, A. (2002). Configural Frequency Analysis: methods, models, and applications. Mahwah, NJ: Lawrence Erlbaum.

von Eye, A. (2005). Comparing tests of multinormality - A Monte Carlo study. InterStat, http://interstat.statjournals.net/ (October).

von Eye, A., & Bogat, G.A. (2004). Testing the assumption of multivariate normality. Psychology Science, 46, 243 - 258.

von Eye, A., & Gardiner, J.C. (2004). Locating deviations from multivariate normality - Concepts and illustrations. Understanding Statistics, 3, 313 - 331.

von Eye, A., & Mun, E.Y. (2003). Characteristics of measures for 2 x 2 tables. Understanding Statistics, 2, 243 - 266.

von Eye, A., & Spiel, C. (1996). Standard and non-standard log-linear symmetry models for measuring change in categorical variables. The American Statistician, 50, 300 - 305.

von Weber, S., von Eye, A., & Lautsch, E. (2004). The Type II error of measures for the analysis of 2 x 2 tables. Understanding Statistics, 3, 259 - 282.

Zhang, P. (1999). Omnibus tests of normality using the Q statistic. Journal of Applied Statistics, 26, 519 - 528.