Comparing Tests of Multinormality under Sparse Data Conditions - a Monte Carlo Study

Alexander von Eye[1]
Michigan State University

[1]

The author is indebted to G. Anne Bogat for helpful comments on an earlier version of this article.
Address correspondence concerning this article to Alexander von Eye, Michigan State University, 107 D Psychology Building, East Lansing, MI 48824-1117, USA; e-mail: voneye@msu.edu

Comparing Tests of Multinormality under Sparse Data Conditions - a Monte Carlo Study

abstract

Mardia's tests of multivariate skewness and kurtosis and von Eye and Gardiner's and von Eye and Bogat's sector and overall tests of multinormality are compared under sparse data conditions. A Monte Carlo study is reported in which five factors were varied: sample size, number of variables, type of distribution (normal, uniform, log-transformed, inverse Laplace-transformed, and cube root-transformed), magnitude of correlation among variables, and the number of segments used for the χ^2-tests. Results suggest that, even under sparse data conditions, (1) the Mardia tests are differentially sensitive to the violations they were designed to detect; (2) the new sector and overall tests are sensitive to all violations included in the simulations; (3) the effects of the small samples can be seen in an increased random component of the data structure; and (4) although the overall and the sector tests are still sensitive to a wide range of violations, the test statistics are not always distributed as χ^2, due to the well known inflation of X^2. These results replicate, in part, the results of von Eye's (2005) study, in which larger samples had been used. However, they also show that sufficiently large samples are needed for valid statistical decisions about multinormality.

Comparing Tests of Multinormality under Sparse Data Conditions - a Monte Carlo Study

From the many tests of multinormality (cf. Romeu & Ozturk, 1993, 1996), four were selected for a study on their performance under sparse data conditions. These four tests include first two specialized tests, sensitive to only one type of violation of multinormality, specifically Mardia's (1970, 1980) well known tests of multivariate skewness and kurtosis. Multinormal distributions show no skew and no excessive kurtosis. Therefore, distributions with skewness and kurtosis which significantly deviate from expectancy cannot be normal. The third test included here is an omnibus test of multivariate normality that was recently proposed by von Eye and Gardiner (2004) and von Eye and Bogat (2004). In an earlier study (von Eye, 2005), this test was shown to be sensitive to a wide range of violations of multinormality. The fourth test included in the comparison is a sector test, also proposed in the articles of von Eye and Gardiner (2004) and von Eye and Bogat (2004). The test examines sectors of the multivariate space, and asks whether the number of cases found in a sector deviates significantly from the number expected based on the assumption of multinormality.

1. Sparse data and tests of multinormality

Sparseness in data is defined by sample sizes that are small relative to the number of variables. In contingency table analysis, tables are called *sparse* when cell counts are small (Agresti, 2002). In continuous data analysis, sparseness results, in a fashion analogous to zero cell counts, in "clumpy" data, that is, data density centers with empty sectors of data space between them.

Sparseness can be a threat to the trustworthiness, even the existence of parameter estimates. For example, based on Haberman's (1973a, b) results, Agresti (2002, p. 393) notes that, in log-linear/logit models, parameter estimates do not exist when *any* cell count is zero. The estimates are finite only when *all* cell counts are greater than zero. When Haberman's extended ML estimators are used, cell frequencies can be estimated to be zero, and log-linear parameter estimates can be infinite. Agresti also notes that, even when a parameter estimate is infinite, this may not be fatal to data analysis. However, this situation can be troublesome, from a user's perspective, because many software packages may not notice this. Instead, they report convergence and the estimated standard errors are extremely large and numerically unstable. The user thus may not realize that the true estimated effect is infinite. The reported effects may then be unstable and statistical decisions may be invalid. In continuous data analysis, the situation is often parallel. In fact, Pendse (2001) states that multidimensional data are *always sparse*.

When data are sparse, application of multivariate procedures of analysis can be risky. In particular, and in addition to the numerical problems discussed by Haberman and Agresti, multivariate procedures that rely on the assumption that data were drawn from multivariate normal populations may produce biased estimates. This may be hard to prevent based on tests of multinormality, because they may not have enough power to detect deviations from multinormality or, worse, may change their sensitivity to data characteristics.

In this article, we report a Monte Carlo study in which sparse multivariate data were created and transformed to reflect various types of deviations from multinormality; tests of multinormality were then performed. Of the many of tests that are available, Mardia's tests of multivariate skewness and kurtosis and von Eye et al.'s overall and sector tests were used. The first two are specialized, and sensitive in particular to deviations from multinormality that result in skewness or kurtosis. These tests are known to perform well for regular size samples (Romeu & Ozturk, 1993; von Eye, 2005; von Eye, Bogat, & von Eye, 2005). The overall test is, in contrast to Mardia's tests, omnibus to a many types of violations. It also performed well for regular size samples (von Eye, 2005; von Eye et al., 2005). The sector test allows one to examine individual cells, and to determine where in the multivariate data space deviations from multinormality are most rampant. This test can thus be seen as bridging overall tests and tests that search for individual multivariate outliers. The sector test also performed well under the conditions simulated thus far (von Eye, 2005; von Eye et al., 2005). In the following paragraphs, we briefly review these four tests (for more detail, see von Eye, 2005, or von Eye & Gardiner, 2005).

<u>Mardia's test of multivariate skewness</u>. Mardia (1970) defined a measure of multivariate skewness as

$$b_{1d} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} r_{ij}^3,$$

where r_{ij}^3 is known as the *Mahalanobis angle* (Mardia, 1980, p. 310) between the vectors $x_i - \bar{x}$

and $x_j - \bar{x}$. The limiting distribution of $Nb_{1d} / 6$ is a χ^2 distribution with $df = d(d + 1)(d + 2)/6$.

Mardia's test of multivariate kurtosis. Mardia's measure of multivariate kurtosis is

$$b_{2d} = \frac{1}{N} \sum_{i=1}^{N} r_i^4,$$

where r_i^4 is the *Mahalanobis distance* of case x_i from its mean, $\bar{x}$. The limiting distribution of

$$\sqrt{N} \; \frac{(b_{2d} - d(d + 2))}{\sqrt{8d(d + 2)}}$$

is normal. For the calculation of both b_{1d} and b_{2d}, the cases under study are assumed to be *iid*.

The sector test. von Eye and Gardiner (2004) and von Eye and Bogat (2004) proposed a test that allows one to examine specified sectors of a multinormal distribution. This test can be viewed as the multivariate extension of the well known χ^2 test of univariate normality. Suppose d variables are analyzed, the range of each of which is split in c_l segments, with $l = 1, ..., d$. Cross-classifying the segmented d variables yields $\prod_{l=1}^{d} c_l$ sectors. The probability for an element to be located in the sector

that is bounded by z_i^1 and z_{i+1}^1 on variable 1, z_j^2 and z_{j+1}^2 on variable 2, ..., and z_k^d and z_{k+1}^d on variable d, where the subscripts index the segments and the superscripts index the variables, is

$$p(z_i^1 - z_{i+1}^1,\, z_j^2 - z_{j+1}^2,\, ...,\, z_k^d - z_{k+1}^d) = \int_{z_i^1}^{z_{i+1}^1} \int_{z_j^2}^{z_{j+1}^2} \cdots \int_{z_k^d}^{z_{k+1}^d} \Psi(z^1, z^2, ..., z^d) \;\; dz^1\, dz^2 \, ... \, dz^d$$

where Ψ indicates the area of the normal distribution. We abbreviate this probability with $p_{i,j,...,k}$. Genz (1992) presented a computational solution for this equation (cf. Gupta, 1963). The sectors defined this way are multivariate-rectangular.

The expected frequency of objects in Sector $s_{i,j,...,k}$ is $e_{i,j,...,k} = Np_{i,j,...,k}$. To answer the question whether Sector $s_{i,j,...,k}$ shows a violation of multinormality, one compares the observed frequency, $o_{i,j,...,k}$, with the expected frequency, $e_{i,j,...,k}$, under the null hypothesis $E[o_{i,j,...,k}] = e_{i,j,...,k}$. If this comparison suggests that Sector $s_{i,j,...,k}$ contains significantly more or fewer objects than expected based on the assumption of a joint normal function of the d variables under study, this sector shows a violation of multivariate normality.

Many tests can be used to examine the null hypothesis for each sector (for an overview, see von Eye, 2002). A prime candidate is the Pearson X^2-component,

$$X^2 = \frac{(o_{i,j,...,k} - e_{i,j,...,k})^2}{e_{i,j,...,k}}$$

with 1 degree of freedom. This statistic is also known as the *Pearson residual* (Agresti, 2002) and the square of the *standardized residual*.

Overall test. As an overall test, Pearson's X^2 can be used. Summing the Pearson residuals yields the test statistic

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

with $df = \left(\prod_{l=1}^{d} c_l \right) - 2d - d_{cov} - 1$, where c_l is the number of segments of the lth variable, and the term d_{cov} indicates the number of correlations (or covariances) taken into account. Usually, $d_{cov} = \binom{d}{2}$, that is, all covariances are taken into account. This statistic can be used as an *overall test of multivariate normality*. Existing results (von Eye, 2005; von Eye et al., 2005) show that this test is omnibus to a broad range of violations of multinormality, including skewness, kurtosis, and symmetry violations.

2. The simulation study

In this section, we describe the simulations that were carried out to study the behavior of the four tests reviewed in the last section, under conditions of sparseness. To perform the simulations, a FORTRAN 90 program was written that varied six factors. The first factor of the simulation was Type of Distribution. Specifically, the following four multivariate distributions were created (cf. Fleishman, 1978; Vale & Maurelli, 1983; von Eye et al., 2005).

(1) Normal distribution. The generator GASDEV from the Numerical Recipes FORTRAN

collection (Press, Flannery, Teukolsky, & Vetterling, 1989) was used to create N(0;1)-distributed data (see also Sicking, 1994). The generator is based on the function RAN1, also provided in the recipe collection. RAN1 returns Gaussian pseudo random deviates. A required, user-specified seed was created using the library MSFLIB that is available in the MS Fortran Power Station. The data created using GASDEV were expected not to deviate from multinormality.

(2) Uniform distribution. The generator RANDOM, available in the Power Station's PortLib function pool, was used to create 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 data created using RANDOM were expected to deviate from multinormality. They are symmetric, but they show kurtosis characterized by heavy tails.

(3) Log-transformed distribution. Uniform variates x that were created as described under (2) were subjected to the logarithmic transformation log(x). The resulting data were expected to exhibit some skewness and elevated kurtosis.

(4) 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 a kurtosis of 0. For $\beta = 1$ and x centered scores, the probability distribution becomes

$$f(x) = \frac{\exp^{-|x|}}{2}.$$

The Laplace distribution is unimodal and symmetric.

A uniform distribution has no skew but exhibits increased kurtosis. Performing an inverse Laplace transformation on a uniform distribution should, therefore, result in a distribution with reduced kurtosis and possibly elevated skewness. However, the Laplace function has no inverse. Therefore, the transformation proposed by von Eye et al. (2005) was performed.

(5) Cube root transformation. This transformation was used to create $y = ½\, x^{1/3}$ 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 both elevated skewness and

elevated kurtosis.

For the simulation, the following data sets were thus created: (1) normally distributed variates; (2) uniformly distributed variates, ranging from 0 to 1; (3) logarithmically-transformed variates (from a uniform distribution) (4) uniformly distributed variates that were subjected to the substitute of the inverse Laplace transformation; and (5) 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:

(ii) The *sample size* varied from 20 to 80, in steps of 20. Thus, 4 different sample sizes were used.

(iii) The *number of segments* of each variable varied from 3 to 5, in steps of 1. Thus, 3 different numbers of segments were used. The number of segments was always the same for all variables.

(iv) The *correlations* among variables. Specifically, variate x_{j+1} was correlated with variate x_j by $x_{j+1} = 0.5 \cdot x_j + x_{j+1} \cdot \rho$. The correlation ρ assumed the four values 0, 0.1, 0.2, and 0.3.

(v) The *number of variables* varied from 3 to 5, in steps of 1. Results for 2 variables are reported separately, in Section 4.

The table with the highest degree of sparseness thus had $5^5 = 3125$ cells for 20 cases. The table with the lowest degree of sparseness had $3^3 = 27$ cells for 80 cases.

The resulting design had 5 (TRANSFORM; type of distribution) x 4 (N; sample size) x 3 (NCAT; number of segments) x 4 (RHO; magnitude of correlations) x 3 (D; number of variables) = 720 different conditions. For the two Mardia tests and for the overall X^2-test, it was determined for each data set whether the test indicated a significant deviation from multinormality (yes = 1; no = 0). For the sector test, the number of sectors with frequencies that were discrepant from those predicted based on the hypothesis of multinormality was counted for each data set.

In the following sections, we ask, for the thus created sparse data situations, (1) whether the four tests under study detected any deviations from multinormality in the two- and higher-dimensional tables; (2) whether the detection of violations of multinormality depends on distribution type; (3) whether the detection of violations of multinormality depends on sample size; (4) whether the detection of violations of multinormality depends on the number of segments in which a data set is partitioned; (5) whether the detection of violations of multinormality depends on the magnitude of correlations between variables; and (6) whether the detection of violations of multinormality depends on the number of variables involved. In addition, we examine the pairwise interactions of the factors of the simulation.

3. Results

To answer the questions raised above, and to make the report in this article parallel to the one in von Eye (2005), we intended to estimate the same logit model as in the earlier article, for each of the dichotomous dependent variables, that is, Mardia's multivariate skewness, Mardia's multivariate kurtosis, and the overall X^2-test. For example, the following model for the dichotomous outcome variable for Mardia's skewness test was originally specified:

$$\begin{aligned} \textit{skewness} = {} & \alpha + \beta_i D + \beta_j N + \beta_k \, TRANSFORM + \beta_l \, RHO \\ & + \beta_{ij} (D \times N) + \beta_{ik} (D \times TRANSFORM) + \beta_{il} (D \times RHO) \\ & + \beta_{jk} (N \times TRANSFORM) + \beta_{jl} (N \times RHO) \\ & + \beta_{kl} (TRANSFORM \times RHO), \end{aligned}$$

where α is the constant, and the β are the regression parameters. The number of categories used for segmentation cannot have an effect on skewness and kurtosis. Therefore, this variable was not used for the analysis of skewness and kurtosis. The first order interactions were included to be able to qualify the main effects, if needed. With the exception of *N* which was used as a 1 *df* variable, all other variables were defined as categorical. Dummy coding was used, and SYSTAT 11's quasi ML method was used for estimation. For kurtosis as dependent variable, the model was intended to be used unchanged. For X^2, the number of segments variable was included in the model (see Section 3.3). For the dependent variable Number of Significant Sectors (NTYPES), the same effects as for X^2 were planned to be estimated, using an ANOVA model.

Before reporting the details of the logit analyses, we note that the dichotomous outcome variables correlated at a moderate to high level. Specifically, we calculate the following simple matching correlations: $S4_{\text{skewness- kurtosis}} = 0.61$, $S4_{\text{skewness-Pearson}} = 0.39$, and $S4_{\text{kurtosis-Pearson}} = 0.46$. These values are similar in level to the ones reported in von Eye (2005), $S4_{\text{skewness- kurtosis}} = 0.51$, $S4_{\text{skewness-Pearson}} = 0.58$, and $S4_{\text{kurtosis-Pearson}} = 0.61$.

3.1 Results for Mardia's (1970) test of multivariate skewness

The model that was planned to be estimated did not converge for the reasons discussed above. The number of zero frequencies was 50%. In addition, the zeros were distributed systematically over the cells of the cross-classification. Specifically, all cells for the normal and the uniform distributions were empty. In addition, the subtable for 3 variables and the cube root-transformed distribution is empty. This is illustrated in Figure 1, in which the number of identified skewness violations is depicted for Type of Distribution (TRANSF) by Number of Variables (NVAR).

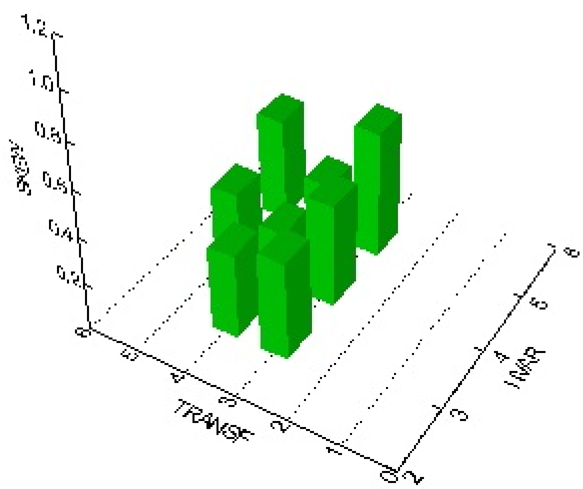


Figure 1: Skewness by type of distribution (TRANSF: 1 = normal distribution, 2 = uniform, 3 = log-transformed, 4 = inverse Laplace-transformed, 5 = cube root-transformed) and Number of Variables (NVAR)

Because of the empty subtables, (i) only three of the five categories of the Type of Distribution variable were used, and (ii) the logit model estimated for skewness was simplified to

$$\begin{aligned} \textit{skewness} = \alpha &+ \beta_i\, D + \beta_j\, N + \beta_k\, \textit{TRANSFORM} + \beta_l\, \textit{RHO} \\ &+ \beta_{ij}\, (D \times N) + \beta_{il}\, (D \times \textit{RHO}) + \beta_{jk}\, (N \times \textit{TRANSFORM}) \\ &+ \beta_{jl}\, (N \times \textit{RHO}) + \beta_{kl}\, (\textit{TRANSFORM} \times \textit{RHO}). \end{aligned}$$

The model was estimated using the quasi ML option in SYSTAT 11's logit module. A total of 28 parameters was estimated, two of which were redundant. In the following paragraphs, we discuss results for the included main effects and interactions, illustrate selected results graphically, and compare results with those from the study with larger samples (von Eye, 2005). McFadden's Rho-squared for the dichotomous outcome variable for Mardia's skewness was 0.21, clearly below the 0.71 found for the same simulation under regular sample size conditions.

Main effects

Number of Variables, D. The first of the two parameters was significant, indicating that the number of identified violations of skewness increased with the number of variables in the study from 3 to 5 variables.

Sample Size. Overall, sample size had no significant effect on the number of skewness violations detected in the simulation.

Distributional Characteristics (TRANSFORM). The effect of distributional characteristics on skewness is illustrated in Figure 1. The figure shows that the portions of violations for the normally distributed and the uniformly distributed random variates are zero. In contrast, the portions of violations for the log-transformed, the inverse Laplace-transformed, and the cube root-transformed variates are large, reaching a maximum of close to 0.5 for the logarithm-transformed distributions. This result differs in two respects from the one reported for regular-size and large samples. First, the number of detected violations is much smaller, suggesting that the α curves are flatter, even when violations are grave. Second, under the present conditions, the logarithmic transformation seems to have the strongest effect.

Magnitude of correlation (RHO). In contrast to the results for larger samples, correlation had no effect.

Interactions

Number of Variables by Sample Size. With no clear pattern in the larger samples study, this interaction is no longer significant.

Number of Variables by Distribution Type. As Figure 1 shows, this interaction would have been fueled mostly by the fact that the subtable "3 variables - cube root transformation" was empty. Therefore, this interaction was not included in the logit model. In addition, this result would have been hard to compare with the one reported in von Eye (2005) because, here, the variable Type of Distribution involves only three transformations.

Number of Variables by Magnitude of Correlation. Although one parameter of this interaction is significant, the (lack of) effect of variable correlation can be considered the same for each number of variables.

Sample Size by Type of Distribution. For the log-transformed and the inverse Laplace-transformed distributions, the portions of detected skewness violations increase with the sample size, in particular for the inverse Laplace-transformed data. For these two distributions, the sample size effect is almost linear. For the cube root-transformed distributions, the sample size effect is erratic, as is illustrated in Figure 2, thus replicating the results reported for larger samples.

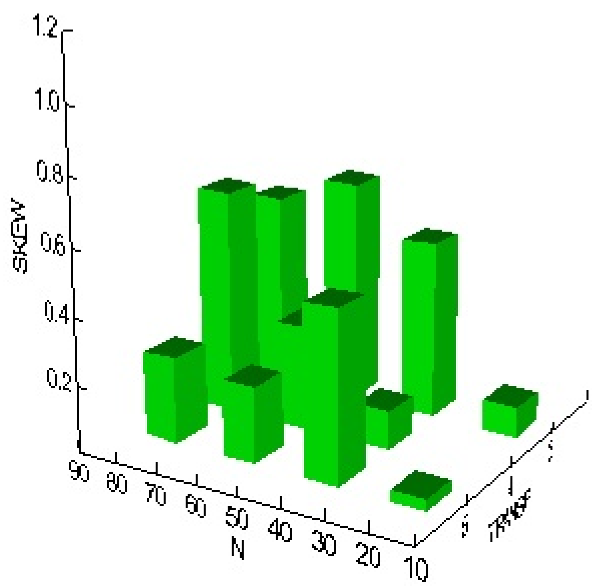


Figure 2: Skewness by type of distribution (TRANSF: 1 = normal distribution, 2 = uniform, 3 = log-transformed, 4 = inverse Laplace-transformed, 5 = cube root-transformed) and Sample Size (N)

Sample Size by Magnitude of Correlation. Not significant.

Type of distribution by Magnitude of Correlation. No effect. The last two results replicate the ones reported for larger sample sizes.

3.2 Results for Mardia's (1970) test of multivariate kurtosis

As for the table for Mardia's skewness test, 50% of the cells in the table for Mardia's multivariate kurtosis test contained zeros. Specifically, there was not a single case in which kurtosis was flagged as extreme for the normally distributed data, and for the inverse Laplace-transformed data for five variables. This is illustrated in Figure 3 in which the number of identified kurtosis violations is depicted for Type of Distribution (TRANSF) by Number of Variables (NVAR).

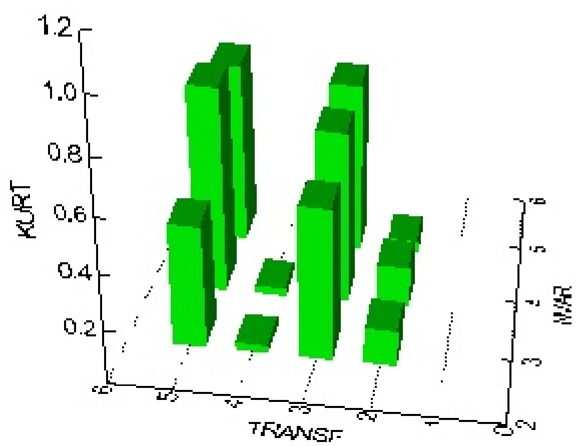


Figure 3: Kurtosis by type of distribution (TRANSF: 1 = normal distribution, 2 = uniform, 3 = log-transformed, 4 = inverse Laplace-transformed, 5 = cube root-transformed) and Number of Variables (NVAR)

Because of the empty subtables, (i) only four of the five categories of the Type of Distribution variable were used, and (ii) the logit model estimated for kurtosis was simplified to

$$kurtosis = \alpha + \beta_i D + \beta_j N + \beta_k TRANSFORM + \beta_l RHO + \beta_{ij} (D \times N) + \beta_{il} (D \times RHO) + \beta_{jl} (N \times RHO) + \beta_{jk} (N \times TRANSFORM)).$$

The model was estimated using the quasi ML option in SYSTAT 11's logit module. A total of 24 parameters was estimated, one of which was redundant. In the following paragraphs, we discuss results for the included main effects and interactions, and illustrate selected results graphically. McFadden's Rho-squared for the dichotomous outcome variable for Mardia's kurtosis was 0.57, clearly below the 0.85 found for the same simulation under regular sample size conditions.

Main effects

Number of Variables, D. No significant effect.

Sample Size, N. No significant effect.

Distributional Characteristics (TRANSFORM). The effect of distributional characteristics on kurtosis is illustrated in Figure 3. The figure shows that the portion of violations for the normally distributed variates is zero. It is only slightly higher for the inverse Laplace-transformed random variates. In contrast, the portions of violations for the uniformly distributed, the log-transformed, and the cube root-transformed variates are larger, approximating or reaching a maximum of over 0.6 for the log-transformed distributions. In the study with larger samples, the normal and the inverse Laplace-transformed distributions also caused the test to flag only small portions of cases. However, the uniform distribution violated the normality assumption about as often as the log-transformed and the cube root-transformed distributions.

Magnitude of Correlation (RHO). No significant effect.

Interactions

the only significant interactions was the one between Type of Distributions (TRANSFORM) and Sample Size (N). It is illustrated in Figure 4. The figure shows that, while for the uniform and the log-transformed distributions, smaller samples lead to fewer significant deviations, for the inverse Laplace-transformed and the cube root-transformed distributions, there is no such decrease.

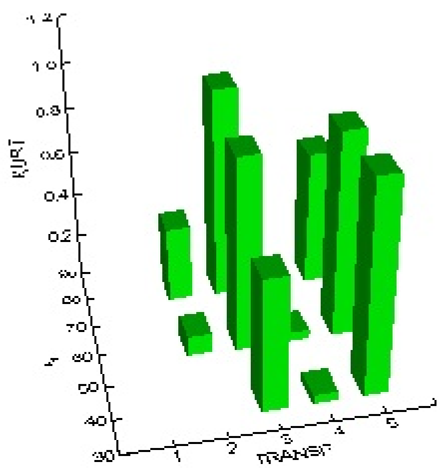


Figure 4: Kurtosis by type of distribution (TRANSF: 1 = normal distribution, 2 = uniform, 3 = log-transformed, 4 = inverse Laplace-transformed, 5 = cube root-transformed) and Sample Size (N)

One more parameter was significant, for the Number of Variables x Sample Size interaction. However, the effect was so small that we refrain from further discussing it.

3.3 Results for overall X^2

The analyses for the overall X^2 were performed using a different model than the analyses for the other two dependent measures, for three reasons. First, the number of segments on each variable (NCAT) and its interactions with the other design variables were also made part of the model. The second reason is that there were no larger subtables empty, as in the models for skewness and kurtosis. However, third, there were a number of subtables with constant cell entries. These were the subtables for the conditions under which the overall X^2 identified all distributions as deviating from multinormality. For the uniform and the inverse Laplace-transformed distributions, not a single table was left unlabeled. For this reason, in order to converge, the model needed to be simplified. Specifically, most interaction terms that included Type of Distributions had to be eliminated from the model, and the uniform and the inverse Laplace-transformed distributions were also excluded from analysis. The logit model specified for the outcome variables for the overall X^2 was, therefore,

$$X^2 = \alpha + \beta_i D + \beta_j N + \beta_k NCAT + \beta_l TRANSFORM + \beta_m RHO + \beta_{ik} (N \times NCAT) + \beta_{jm} (N \times RHO) + \beta_{jm} (NCAT \times RHO) + \beta_{kl} (N \times TRANSFORM) + \beta_{lm} (TRANSFORM \times RHO).$$

McFadden's Rho-squared for the main effect model with Pearson's X^2 as the dependent variable was 0.40, also below the 0.78 that had been estimated for the larger samples. A total of 30 parameters was estimated, including the constant. Two of these parameters were redundant.

Main effects

Number of Variables. The first parameter for this main effect was significant, indicating that the increase in variables from 3 to 4 led to an increase in the number of flagged distributions. However, the increase from 4 to 5 variables was minimal.

Sample Size. No effect. It should be noted, however, that this may be due to a ceiling effect because, for each of the sample sizes (20 through 80, in steps of 20), the portion of detected violations was about 95%. This is below the portion for the larger samples (von Eye, 2005), but still extremely high, and higher than for the measures of skewness and kurtosis.

Number of Variable Categories. No effect.

Type of Distribution. One of the two parameters was significant. The portion of flagged distributions was higher for the log-transformed (95%) than for the normal and the cube root transformed distributions (90% for both).

Magnitude of Correlation. No effect.

Interactions

Sample Size by Number of Segments. There was no sample size effect for 5 segments. For 3 and 4 segments, there was an inverse sample size effect such that the number of flagged distributions decreased as the sample size increased.

Sample Size by Type of Distribution. For the normal and the cube root-transformed distributions, the number of flagged distributions decreased as the sample size increased from 20 to 60, and increased for $N = 80$. For the log-transformed distribution, the number of flagged distributions increased with the sample size.

None of the other interaction parameters was significant.

3.4 Results for the number of sectors indicating multinormality deviations

Again in the attempt to create results that are comparable with the ones of the simulation for larger samples, a linear model was estimated for the number of sectors flagged as significantly deviating from expectancy under a multinormal distribution. This model included the main effects of the variables Number of Variables (NVAR), Sample Size (N), Number of Categories for each variable (NCAT), Type of Distribution (TRANSF), and Magnitude of Correlation (RHO) as well as all pairwise interactions. The multiple R^2 for this model was 0.89, suggesting that a very large portion of variability of the dependent measure can be explained by this model. This value is slightly larger than in the study with larger samples, where it was 0.84. Table 1 displays the ANOVA table for this design (from SYSTAT 11).

Table 1: ANOVA table for analysis of number of deviant sectors

Source	Sum-of-Squares	*df*	Mean-Square	F-ratio	*p*
NVAR	583.839	2	291.919	4.961	0.007
N	26737.700	1	26737.700	454.356	0.000
NCAT	306.072	2	153.036	2.601	0.075
TRANSF	100.352	4	25.088	0.426	0.790
RHO	8.985	3	2.995	0.051	0.985
NVAR*N	23326.234	2	11663.117	198.192	0.000
NCAT*NVAR	17909.939	4	4477.485	76.086	0.000
TRANSF*NVAR	3739.133	8	467.392	7.942	0.000
RHO*NVAR	13.669	6	2.278	0.039	1.000
NCAT*N	16532.507	2	8266.254	140.469	0.000
TRANSF*N	3486.815	4	871.704	14.813	0.000
RHO*N	14.067	3	4.689	0.080	0.971
RHO*TRANSF	122.642	12	10.220	0.174	0.999
RHO*NCAT	9.753	6	1.625	0.028	1.000
TRANSF*NCAT	1402.975	8	175.372	2.980	0.003
Error	38368.523	652	58.847		

The results in Table 1 correspond closely to the ones found for larger samples. The interactions of Magnitude of Correlation (RHO) with Number of Categories (NCAT), Sample Size (N) and with Type of Distribution (TRANSF) are not significant in either analysis, and neither is the main effect of RHO. The only difference to the results of the earlier study is that, now, the main effect of Type of Distribution and Number of Categories are no longer significant. In the following sections, we interpret and, selectively, illustrate the significant (and some of the non-significant) effects listed in Table 1.

Main effects

Number of Variables (NVAR). The number of identified deviant sectors increases linearly with the number of variables.

Sampe Size (N). The number of identified deviant sectors increases linearly with the sample size.

Number of Segments (NCAT). The number of identified deviant sectors increases linearly with the number of segments for each variable. Although clearly visible in a bar chart (not shown here), this effect is not significant.

Type of Distribution (TRANSF). Although, as in the simulation with the larger samples, the uniform and the inverse Laplace-transformed distributions yielded the largest numbers of deviant sectors, this main effect was not significant. In contrast to the earlier study, the normal and the cube root-transformed distributions did yield numbers of deviant sectors only slightly below those for the uniform distribution.

Interactions

Number of Variables (NVAR) by *Sample Size* (N). Figure 6 shows that there is no sample size effect for 3 variables. For 4 and 5 variables, there is a clear effect.

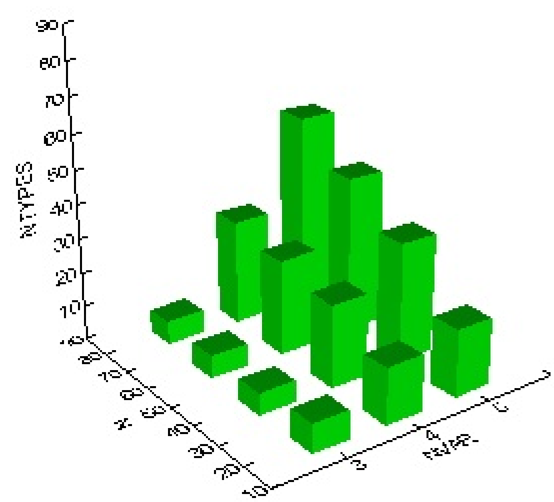


Figure 6: Number of Variables (NVAR) by Sample Size (N) interaction for Number of deviant sectors (NTYPES)

Number of Variables (NVAR) by *Number of Segments* (NCAT). The number of identified deviant sectors increases with both the number of variables and the number of segments. However, the Number of Segments effect is weakest for 3 variables.

Number of Variables (NVAR) by *Type of Distribution* (TRANSF). The increase in the number of deviant sectors is weakest for the log-transformed distribution. This effect is, overall, rather weak.

Number of Segments (NCAT) by *Sample Size* (N). The sample size effect exists only for 3 or more segments.

Sample Size (N) by *Type of Distribution* (TRANSF). The sample size effect is stronger for the uniform and the inverse Laplace-transformed distributions than for the other three distributions.

Type of Distribution (TRANSF) by *Number of Segments* (NCAT). The number of segments effect is strongest for the normal and the cube root-transformed distributions. This effect is rather weak.

3.5 Interpretation and discussion of results

Before discussing the results presented in the last three sections, we summarize the current results in Table 2.

Not a single main effect was significant for all four tests in this simulation. Two of the interactions with Type of Distribution, the one with the number of variables and the one with sample size were significant for all four tests. In each case, the number of detected deviations from multinormality showed no sample size effect or number of variables effect for the normally distributed data. For Mardia's tests, the number of flagged violations was zero. For the overall X^2 test, there was a sample size effect for number of variables, but none for sample size. In contrast, Mardia's tests showed sample size and number of variables effects for the log-transformed, the inverse Laplace-transformed, and the cube root transformed distributions. For the X^2 test, the effects of the cube root-transformed distribution were parallel to those for the normal distribution. For the remaining three distributions, there were practically no sample size or number of variables effects, possibly due to a ceiling effect.

In regard to data sparseness, we note that Mardia's tests respond strongly. For the sample size/number of variables patterns studied here, skewness is not found at all for the normal and the uniform distributions. So, the α curve is not only flat, it is glued to zero. The same applies to kurtosis for normally distributed data. For the other simulated distributions, sparseness effects are as described above. The effects of sparseness, however, are not solely smaller numbers of identified violations. Rather, there is an increased portion of erratic behavior of Mardia's tests, as compared to the behavior observed for larger samples. This is reflected in the lower R^2 equivalents in the logit models.

Table 2: Summary of simulation results (symbols: ✔ = effect significant; - : parameter not estimated; 0: null hypothesis prevails)

	Test			
Effect	Skewness[a]	Kurtosis[b]	Overall X^2	Sector [c]
D	✔	0	✔	✔
N	0	0	✔	✔
NCAT	-	-	✔	0
TRANSF	✔	✔	✔	0
RHO	0	0	✔	0
D x N	0	✔	✔	✔
D x NCAT	-	-	✔	✔
D x TRANSF	✔	✔	✔	✔
D x RHO	0	0	✔	0
N x NCAT	-	-	✔	✔
N x TRANSF	✔	✔	✔	✔
N x RHO	0	0	0	0
NCAT x TRANSF	-	-	✔	✔
NCAT x RHO	-	-	✔	0
TRANSF x RHO	0	✔	✔	0

[a] Interaction D x TRANSF excluded due to identification problems; normal and uniform distributions not taken into account

[b] Interaction D x TRANSF excluded due to identification problems; normal distribution not taken into account

[c] ANOVA model estimated

One surprising result is that the sector and the overall X^2 tests seem to have colossal power, even under conditions of sparseness. This result seems to contradict the fact that X^2 is sample size dependent. We therefore added an exploratory step to our analyses, aimed at studying the behavior of the overall X^2 under additional sparseness conditions. The results of this step are presented in the next section.

4. The overall X^2 and sector tests under sparseness, in square tables

In this section, we study the behavior of the overall X^2 and the sector test under two conditions. The first parallels the one simulated in Section 3. However, instead of asking questions concerning the number of distributions identified as violating multinormality, we now ask whether the X^2 values are distributed as χ^2 under their table- and model-specific degrees of freedom. Second, for illustrative purposes and to complement the results for three and more dimensions reported in Section 3, we perform a simulation on square tables, and ask questions concerning the comparative sensitivity of X^2. The case of square tables is important for a number of reasons. First, square tables are of interest in many areas of social science research, for example rater agreement (von Eye & Mun, 2005). Second, when square tables are small, there may be cases in which even sparseness will not render the X^2 tests invalid. Third, in square tables, there is a relationship between multinormality and axial symmetry (von Eye et al., 2005). Fourth, from a didactical perspective, it is clear that square tables are easy to depict.

4.1 Approximating the χ^2 distribution under sparseness

It is well known that Pearson's X^2 is best applied when the expected cell frequencies are larger than a certain minimum. Under most conditions, the statistic will approximate the χ^2 distribution well if the expected cell frequency is $\hat{m} \geq 1.0$ (for more detailed discussions, see Larntz, 1978; Koehler & Larntz, 1980; von Eye, 2002). Under conditions of sparseness, the estimated expected cell frequencies can be very small. For example, expected cell frequencies can be estimated to be less than 0.01. If this happens, Pearson's X^2 will explode, and be unrelated to the χ^2 distribution. For example, if the observed cell frequency is only 1, we obtain, for the expected frequency of 0.01,

$X^2 = \frac{(1 - 0.01)^2}{0.01} = \frac{0.99^2}{0.01} = 98.01$. When a table is very large and very sparse, even more extreme situations can occur.

In the simulations reported in Section 3, situations of the kind considered here were very likely to happen. Consider the table with 5 variables, a sample of 20, and 3125 cells. If, in this table, we observe marginal frequencies of 1 for each of the variables, the estimated expected cell frequency for the cell at the intersection of these marginals is $\frac{1^5}{20^4} = 0.00000625$. For this cell, we calculate a Pearson's X^2 of 159,998. Therefore, it will be almost impossible to obtain a Pearson's $X^2 = 0.0$ or a Pearson's X^2 close to its expectancy, the *df*. For example, for 3 variables and three segments per variable, that is, for 27 cells, we have a situation with $df = 17$ (main effect model; correlations among the three variables taken into account), the simulation yielded a minimum Pearson's X^2 value of 14.74, and a maximum of 125.511. The Q-plot for the present situation appears in Figure 7.

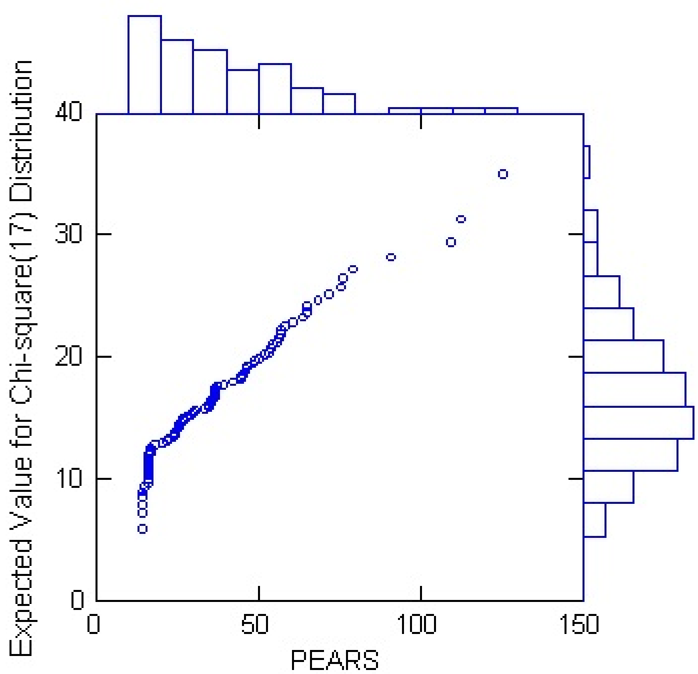


Figure 7: Q-plot of simulated Pearson's X^2 values for samples of size N ≤ 80, 3 variables with 3 categories each (27 cells)

Figure 7 shows that the distribution of the statistic does not approximate the χ^2 distribution well. There are no scores near zero, the empirical distribution is far more skewed than the theoretical one, and the right hand tail is far too heavy and long. In the present simulations, even more extreme situations occurred, as the Pearson X^2 ranged between 14.74 and 23,070.78, for all conditions. In all, the number of tables with extremely small expected cell frequencies was close to 100%.

Accordingly, the sector test results cannot be trusted either. If an expected cell frequency in any table is too small, the sector test will label this sector as deviating even if the difference between the observed and the expected frequencies is less than 1.0, if the observed frequency is 1 or larger. So, it is not surprising that, under some of the simulated conditions, the number of cells that are labeled as deviant equals the number of cells in the table (see Figure 7, above).

We thus conclude once more that, unless each estimated expected cell frequency is larger than a minimum that is above 0.5, the Pearson X^2 cannot be trusted to approximate the χ^2 distribution. Accordingly, the overall X^2 test and the sector test cannot be used to test hypotheses concerning multinormality if one or more expected cell frequencies are below this minimum. The high rates of sectors and distributions identified as deviating from multinormality are due to inflated X^2

4.2 Comparing the sensitivity and specificity of Mardia's tests of skewness and kurtosis and the overall and sector Pearson X^2 tests

In an earlier study (von Eye et al., 2005), it was shown that the overall X^2 and sector tests were sensitive to violations of symmetry in square tables with regular size cell frequencies. In this section, we ask, for square but sparse tables, the following four questions:

(1) Are Mardia's tests of multivariate skewness and kurtosis sensitive to violations of bivariate normality?

(2) Are von Eye et al.'s overall X^2 and sector tests sensitive to violations of bivariate normality?

(3) Are the four tests differentially sensitive to violations of bivariate normality?

(4) Do von Eye et al.'s test statistics approximate a χ^2 distribution?

A separate simulation was performed for square tables. The first factor varied in this part of the study was Type of Distribution (as in Section 3). In addition, the following data characteristics were varied in the simulation:

(ii) The *cell size* varied from 1 to 16, in steps of 1. Thus, 16 different sample sizes were used.

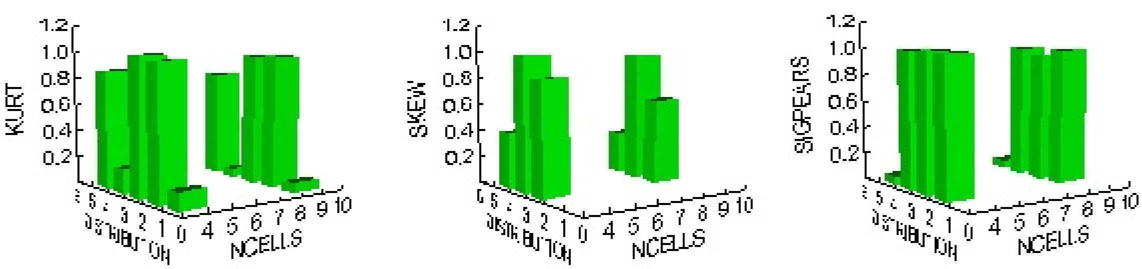
(iii) The *number of segments* of each variable: 2 and 3 segments were used. The number of segments was always the same for the two variables of the table.

(iv) The *correlations* between variables. As in Section 3, variate x_{j+1} was correlated with variate x_j by $x_{j+1} = 0.5 \cdot x_j + x_{j+1} \cdot \rho$. The correlation ρ assumed the five values 0, 0.1, 0.2, 0.3, and 0.4.

The table with the highest degree of sparseness thus had $3^2 = 9$ cells for 4 cases. The table with the lowest degree of sparseness had $2^2 = 4$ cells for 64 cases.

The resulting design was thus a 5 (TRANSFORM; type of distribution) x 16 (N; cell size) x 2 (NSEG; number of segments) x 5 (RHO; magnitude of correlations) design with 800 different conditions. For the two Mardia tests and for the overall X^2-test, it was determined for each data set whether the test indicated a significant deviation from multinormality (yes = 1; no = 0). For the sector test, the number of sectors with frequencies that were discrepant from those predicted based on the hypothesis of multinormality was counted for each data set.

Multinormality violations in square tables. To answer the first two questions, we inspect Figure 8. The three panels of Figure 8 depict, from left to right, the portions of cells flagged as deviant by Mardia's tests of kurtosis and skewness, and by the overall X^2 test.



I ss and

l: 1 =

place-

The panel for kurtosis shows that the results from the simulation with the larger and higher-dimensional tables (see Figure 3) are largely replicated. Mardia's kurtosis test is sensitive to the multinormality violations caused by the uniform, the log-transformed, and the cube root-transformed, and, to a lesser extent, the inverse Laplace-transformed distributions. The middle panel shows a replication of the results for skewness. Mardia's test is sensitive to violations of multinormality that are caused by the logarithmic, the inverse Laplace, and the cube root transformations (cf. Figures 1 and 2).

The results for the overall X^2 test also replicate the ones from the simulation with the larger and higher-dimensional tables. We see that the normal and the cube root-transformed distributions were either not flagged at all (normal distribution) or only very rarely. In contrast, almost all other distributions from the three other transformations were flagged.

However, these results also differ from the ones reported in Section 3. Specifically, the number of transformations that did not cause any distribution to be flagged is smaller here. Compare, for example, the left panel of Figure 8 with Figure 3. In Figure 8, kurtosis violations are noted for each transformation. In the simulation with the larger and higher-dimensional tables, the normal distribution was never flagged. The main reason for this difference is most likely that the present tables were less extremely sparse than the ones used in Section 3.

To conclude, the first two questions asked in this section can be answered in the affirmative. Given the results from the earlier studies, any other result would have been shocking. The most important result produced in this section thus far is that, when there are no strong violations of multinormality, the overall X^2 test will not be inflated (see normal and cube root-transformed distributions; Figure 8). Logically, the same applies to the Sector test (not illustrated here). When we answer Question 4, we will see whether the inflation occurs in square tables when multinormality is violated. Before we do this, however, we ask whether the four tests are differentially sensitive to violations of multinormality.

Differential sensitivity of tests of multinormality. To study the differential sensitivity of the two Mardia tests and the overall X^2 test, we first make an overview statement. Second, we qualify this statement, if needed, based on the distributions used in the present simulations. To create the overview statement, we cross the three dichotomous variables used as dependent measures in the above logit models. The resulting 3-way table is then analyzed using Configural Frequency Analysis (CFA; Lienert & Krauth, 1975; von Eye & Gutiérrez Peña, 2004). CFA tells us whether patterns of declaring distributions as deviating from multinormality occur more often than expected under the assumption of independence of the three tests (types), less often (antitypes), or as often as expected. Table 3 displays the CFA results. In the table, 0 indicates that a distribution was not labeled as deviating; 1 indicates that violations were detected. The order of variables in the cross-classification is overall Pearson X^2 (P), skewness (S), and kurtosis (K). The z test was employed. The Bonferroni–protected significance threshold was α* = 0.00625.

Table 3: CFA of the cross-classification of the results from the overall Pearson X^2 test (P) and the two Mardia tests of multivariate skewness (S) and kurtosis (K)

Configuration PSK	m	$\hat{m}$	z	P	Type/ Antitype?
000	169	75.434	10.773	.00000	Type
001	98	111.979	-1.321	.09325	
010	7	52.964	-6.316	.00000	Antitype
011	45	78.624	-3.792	.00007	Antitype
100	7	113.741	-10.009	.00000	Antitype
101	196	168.846	2.090	.01832	
110	139	79.861	6.618	.00000	Type
111	139	118.551	1.878	.03019	

The first type in Table 3 indicates that the three tests jointly indicate, more often than compatible with the assumption of independence, that distributions are not in violation of multinormality. The following two antitypes suggest that it is less likely than expected that the skewness of a distribution is extreme and this is not flagged by the overall X^2 test, regardless of what Mardia's kurtosis test indicates. Accordingly, it is also less likely than expected that only the overall X^2 test indicates a violation but neither of the Mardia tests does. The second type suggests that, more often than expected, the overall X^2 test and Mardia's skewness test flag distributions as violating multinormality, while the kurtosis test does not indicate any violation.

These results indicate again that Mardia's tests are specific in their sensitivity to violations of multinormality. These results also suggest that, whenever a violation is flagged by any of the Mardia tests, it will also be flagged by the overall X^2 test. This result confirms the earlier conclusion that the overall X^2 test is omnibus to a large number of violations of multinormality (von Eye et al., 2005).

We now proceed and ask whether the statements made based on the CFA of the frequency distribution in Table 3 can be made crisper by taking the specific distributions realized in the present simulation study into account. We repeat the analyses for Table 3 for a table that includes Type of Distribution as a factor. The results of the CFA of the resulting 5 x 2 x 2 x 2 cross-tabulation appear in Table 4. As for Table 3, the *z*-test was used. The Bonferroni-protected α was α* = 0.00125.

Table 4: CFA of the cross-classification of the results from the overall Pearson X^2 test (P), the two Mardia tests of multivariate skewness (S) and kurtosis (K), for Type of Distribution (D)

Configuration DPSK	*m*	$\hat{m}$	*z*	*P*	Type/ Antitype?
1000	145	15.087	33.447	.00000000	Type
1001	15	22.396	-1.563	.05905062	
1010	0	10.593	-3.255	.00056771	Antitype
1011	0	15.725	-3.965	.00003665	Antitype
1100	0	22.748	-4.770	.00000092	Antitype
1101	0	33.769	-5.811	.00000000	Antitype
1110	0	15.972	-3.997	.00003215	Antitype
1111	0	23.710	-4.869	.00000056	Antitype
2000	1	15.087	-3.627	.00014357	Antitype
2001	0	22.396	-4.732	.00000111	Antitype
2010	0	10.593	-3.255	.00056771	Antitype
2011	0	15.725	-3.965	.00003665	Antitype
2100	4	22.748	-3.931	.00004234	Antitype
2101	155	33.769	20.862	.00000000	Type
2110	0	15.972	-3.997	.00003215	Antitype
2111	0	23.710	-4.869	.00000056	Antitype
3000	3	15.087	-3.112	.00092984	Antitype
3001	2	22.396	-4.310	.00000818	Antitype
3010	0	10.593	-3.255	.00056771	Antitype
3011	0	15.725	-3.965	.00003665	Antitype
3100	0	22.748	-4.770	.00000092	Antitype
3101	38	33.769	.728	.23329130	
3110	0	15.972	-3.997	.00003215	Antitype

3111	117	23.710	19.159	.00000000	Type
4000	1	15.087	-3.627	.00014357	Antitype
4001	0	22.396	-4.732	.00000111	Antitype
4010	0	10.593	-3.255	.00056771	Antitype
4011	0	15.725	-3.965	.00003665	Antitype
4100	3	22.748	-4.141	.00001734	Antitype
4101	0	33.769	-5.811	.00000000	Antitype
4110	138	15.972	30.533	.00000000	Type
4111	18	23.710	-1.173	.12045648	
5000	19	15.087	1.008	.15684721	
5001	81	22.396	12.384	.00000000	Type
5010	7	10.593	-1.104	.13481979	
5011	45	15.725	7.383	.00000000	Type
5100	0	22.748	-4.770	.00000092	Antitype
5101	3	33.769	-5.295	.00000006	Antitype
5110	1	15.972	-3.746	.00008976	Antitype
5111	4	23.710	-4.048	.00002586	Antitype

The CFA results in Table 4 show a very large number of types and antitypes. For four of the five distributions, there is a clear structure. The first distribution is the normal. We find a type for configuration 1000, indicating that, for a normal distribution, it is far more likely than expected under the assumption of variable independence that all three tests indicate no violation of multinormality. With only one exception (configuration 1001), all other result patterns are far less likely than expected. This result indicates that, for square tables and normally distributed data, the three tests systematically yield the same results, practically all supporting the null hypothesis.

For the second, the uniform distribution, we obtain a similarly clear pattern of results. There is a type for configuration 2101. It indicates that, whereas Mardia's kurtosis test and the overall Pearson X^2 test flag the uniform distribution as violating the multinormal, Mardia's skewness test, logically, does not indicate any violation. All other configurations constitute antitypes, suggesting that other classification patterns occur less likely than expected. This includes the pattern in which none of the tests notices a violation (configuration 2000). From this result, we can conclude that only Mardia's kurtosis test and the overall Pearson X^2 test are sensitive to kurtosis violations.

The third distribution resulted from a logarithmic transformation. The type, constituted by configuration 3111, shows that the three tests are more likely than expected to jointly identify a distribution as violating multinormality. With only one exception (configuration 3101), all other result patterns constitute antitypes.

For the inverse Laplace-transformed, the third distribution, we obtain an analogous picture. The type, constituted by configuration 4110, suggests that this transformation results, under the present conditions, in distributions with extreme skewness. Mardia's skewness test and the overall Pearson X^2 test flag the distributions as deviant. With the exception of pattern 4111, all other configurations constitute antitypes.

The cube root-transformed distribution, number 5 in the series of distributions, is the exception to the rule. Only here, we find two configurations that constitute types. The first, found for configuration 5001, indicates that there are more distributions than expected that are flagged as deviant by only Mardia's kurtosis test. The second type, found for configuration 5011, indicates that more distributions than expected were flagged by both the overall Pearson X^2 test and Mardia's kurtosis test. Of the remaining configurations, 4 constitute antitpyes, and 2 are non-suspicious.

In all, the CFA of the frequencies of flagged distributions shows that, for sparse tables,

1. whenever either of the Mardia tests suggests that a distribution violates multinormality, the overall Pearson X^2 test will suggest the same conclusion (the only noteworthy exception to this rule is found for configuration 5001);
2. The three tests under study suggest the same decision only when distributions are multinormal; and
3. When there are clearly either only skewness or kurtosis violations, the two Mardia tests practically never suggest the same decision (see types 2101 and 4110);
4. The overall X^2 test is thus omnibus to violations of skewness and kurtosis and, as we know from earlier results (von Eye et al., 2005), symmetry.

Considering the inflation of the X^2 statistics that was so obvious in the extremely sparse three- and higher-dimensional tables studied in Section 3 of this article, we now ask whether the Pearson X^2 statistic is distributed more closely to the χ^2 under the present conditions, that is, when sparseness is less extreme, in two-dimensional tables.

Approximation characteristics of X^2. The above results show consistent and expected results. However, as was found for the higher-dimensional tables, the X^2 statistic can be considerably inflated when expected cell frequencies are small. Here, we ask whether this inflation which can be expected to occur in two-dimensional tables also, prevents the X^2 statistic from being distributed as χ^2.

We present Q-plots for two cases. The first case includes the distributions with cell sizes 1 through 8. The second includes the distributions with cell sizes 9 through 16. Figure 9 displays the two Q-plots. Please note that the degrees of freedom are $df = 1$ both for 2 and 3 segments. The reason is that, for 2 segments, the variable correlation was not taken into account, to prevent the model from becoming saturated.

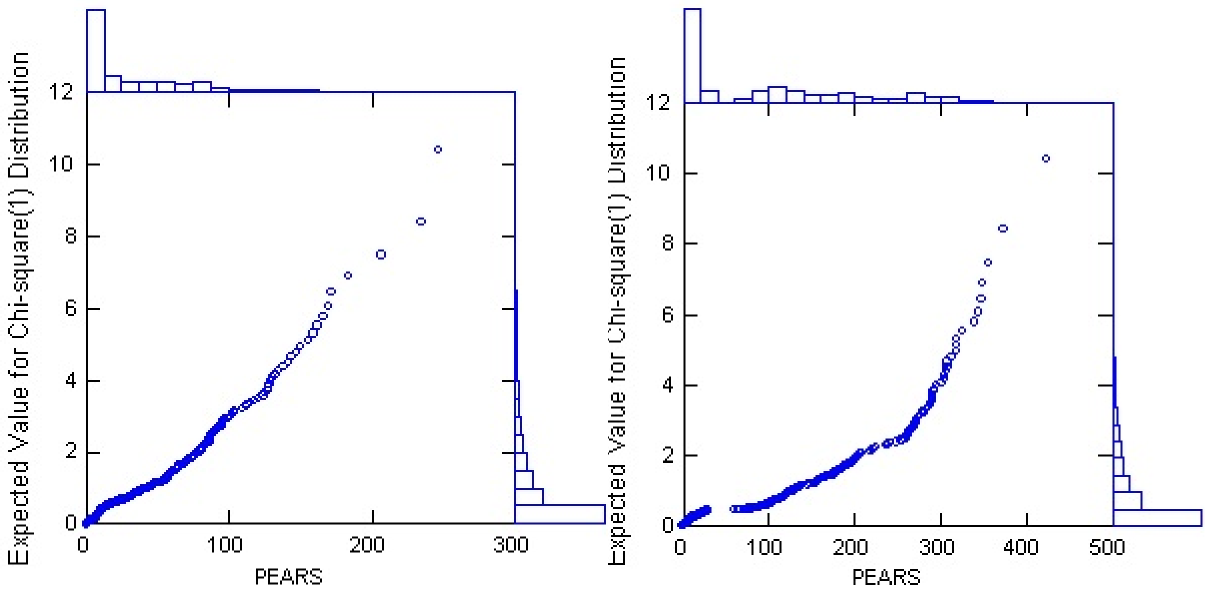


Figure 9: Q-plot for the X^2 statistic for sparse square tables for cell sizes 1 through 8 (left panel) and 9 through 16 (right panel)

Both panels in Figure 9 show the expected result. The X^2 statistic of the overall test of multinormality is extremely heavy-tailed. The sample size affects the distribution such that inflated scores become both more likely and more extreme.

5. Discussion

In this study, we simulated sparse data situations with the goal of examining the performance of tests of multinormality. The four tests under study were Mardia's (1970, 1980) tests of multivariate skewness and kurtosis and von Eye et al.'s (von Eye & Bogat, 2004; von Eye & Gardiner, 2004) sector and overall X^2 tests of multinormality. Because the behavior of Mardia's tests has been studied before (Lee, 1998; Mecklin & Mundfrom, 2003, 2005), the present study focused on the new tests that were proposed by von Eye and collaborators.

The results of the simulations suggest an interesting pattern. First, under sparse data conditions, all of the tests are sensitive to the same data characteristics as under regular data conditions. Specifically, Mardia's tests are sensitive to excessive skewness and kurtosis, respectively, and the sector and overall X^2 tests are omnibus to these and other violations such as symmetry. However, the X^2 tests also, and strongly, show the well-known inflation when an estimated expected cell frequency is small and the corresponding observed frequency is greater than zero.

It is surprising to see that, even under extreme sparseness, the statistics are still sensitive to violations of multinormality. However, this characteristic is watered down by the inflation effect. Therefore, we conclude, for sparse data situations:

1. If the X^2 statistic of the overall test of multinormality suggests that the null hypothesis of multinormality can be retained, (a) the statistic can be trusted, (b) is close to its sampling distribution, and (c) it can be assumed that the distribution at hand was drawn from a multinormal population. However, statistical power may be an issue.
2. If the X^2 statistic of the overall test of multinormality suggests that the null hypothesis of multinormality must be rejected, two cases must be distinguished. In the first, inflation of X^2 took place. This can easily be identified by finding cells with expected frequencies less than 0.5 and observed frequencies of 1 or larger. In this case, the test statistic cannot be trusted, and the decision concerning multinormality must be based on other tests, e.g., the two Mardia tests. In the second case, there are no small expected cell frequencies in tandem with observed frequencies of 1 or larger. In this case, it can safely be assumed that multinormality does not exist. However, even in this case, we caution against trusting the tail probabilities that can be calculated for the Pearson test statistic. The reason for this caution is that the statistic is not distributed as χ^2. Instead of approximating the χ^2 distribution, the distribution of the test statistic becomes irregular or, for $df = 1$, almost rectangular.

So, what is a warm body to do when the X^2 statistics of the multinormality tests are inflated and sectors of strong deviation are searched for? Fortunately, there is hope. It is possible to reduce sparseness while still testing the hypothesis that a multivariate distribution is multinormal. Based on the fact that a multivariate distribution cannot be multinormal if only one of the variables that constitute the distribution is non-normal, the following ascending strategy can be considered. The strategy results in the identification of variables that are jointly multinormal.

1. Subject all individual variables to tests of normality. Clearly, for any given N, sparseness will be less of a problem when individual variables are examined. To come to a decision concerning the distributional characteristics of a variable, any of a large number of tests can be used, including, for example, the Kolomogoroff-Smirnoff test, skewness- and kurtosis tests, and univariate versions of the X^2 overall and sector tests. Based on the results of these tests, those variables for which the normal distribution hypothesis was rejected can be eliminated.
2. Subject all pairs of the remaining variables to tests of bivariate normality. Here, the Mardia tests and the X^2 overall and sector tests can be used. Eliminate all pairs that are not binormal.
3. Continue the variable selection with all groups of three, four, and more variables, until all sets of multinormally distributed variables are identified (or until sparseness prevents one from making decisions that can be trusted).

References

Agresti, A. (2002). *Categorical data analysis*, 2nd ed. New York: Wiley.

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

Haberman, S.J. (1973a). The analysis of residuals in cross-classification tables. *Biometrics*, *29*, 205 - 220.

Haberman, S.J. (1973b). Log-linear models for frequency data: Sufficient statistics and likelihood equations. *The Annals of Statistics*, *1*, 617 - 632.

Koehler, K.J., & Larntz, K. (1980). An empirical investigation of goodness-of-fit statistics for sparse multinomials. *Journal of the American Statistical Association*, *76*, 336 - 344.

Larntz, K. (1978). Small-sample comparisons of exact levels for chi-squared goodness-of-fit statistics. *Journal of the American Statistical Association*, *75*, 336 - 334.

Lee, Y.-H. (1998). *Fisher information test of normality*. Dissertation, Virginia Polytechnic Institute and State University, Department of Statistics. Blacksburg, VA.

Lienert, G.A., & Krauth, J. (1975). Configural Frequency Analysis as a statistical tool for defining types. *Educational and Psychological Measurement*, *35*, 231 - 238.

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.

Mecklin, C.J., & Mundfrom, D.J. (2003). On using asymptotic critical values in testing for multivariate normality. *InterStat*, http://interstat.stat.vt.edu/InterStat/ ARTICLES/2003/articles /J03001.pdf (January issue).

Mecklin, C.J., & Mundfrom, D.J. (2005). A Monte Carlo comparison of the Type I and Type II error rates of tests of multivariate normality. *Journal of Statistical Computations and Simulation*, *75*, 93 - 107.

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

Pendse, N. (2001). Multidimensional data structures. In *The OLAP Report* http://www.olapreport.com/MDStructures.htm

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.

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

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

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

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

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

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

von Eye, A., Bogat, G.A., & von Eye, M. (in press). Multinormality and symmetry: A comparison of two statistical tests. *Psychology Science*.

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

von Eye, A., & Gutiérrez Peña, E. (2004). Configural Frequency Analysis - the search for extreme cells. *Journal of Applied Statistics*, *31*, 981 - 997.

von Eye, A., & Mun, E.Y. (2005). *Modeling rater agreement - manifest variable approaches*. Mahwah, NJ: Lawrence Erlbaum.