Psychology Science, Volume 46, 2004 (2), p. 243-258

Testing the assumption of multivariate normality

ALEXANDER VON EYE[1], G. ANNE BOGAT

Abstract

Methods of assessing the degree to which multivariate data deviate from multinormality are discussed. The best known of these methods, Mardia's tests of multivariate skewness and kurtosis, allow one to test null hypotheses that are compatible with the assumption of multinormality. However, if these null hypotheses are rejected, researchers do not know whether particular sectors carry inordinate amounts of the violations. A sector test and an omnibus test are proposed. The sector test allows one to identify subspaces that contain different numbers of cases than expected on the assumption of multinormality. The omnibus test allows one to test whether, overall, the hypothesis of a multinormal distribution is tenable. Mardia's tests and the new sector and omnibus tests are applied to data from a project on parenting abilities of adolescents.

Key words: multivariate normal distribution; multivariate skewness and kurtosis; multivariate sector tests; multinormal distribution test

[1] Alexander von Eye, Michigan State University, Department of Psychology, 107D Psychology Building, East Lansing, MI 48824-1116, USA; E-mail: voneye@msu.edu

Statistical hypotheses in such methods as the multivariate analysis of variance (MANOVA) are usually tested under three assumptions:

1) the observations are independently distributed;
2) the observations are normally distributed; and
3) the observations have a common variance-covariance matrix.

The second assumption implies that the observations are randomly drawn from a multivariate normal (multinormal) population. Tests are called *robust* if their significance level (Type I error probability) and its power (1 - Type II error probability) are insensitive to violations of the assumptions based on which these tests are derived (Ito, 1980).

To be able to determine whether a test is applicable and robust, one needs to know whether the sample at hand was drawn from a multinormal distribution. There exists a large number of tests for multinormality. These tests allow one to examine hypotheses that are consistent with the assumption of multinormality. They do not allow one to test the distribution directly. In this article, we review a selection of existing tests, and present and apply a new test of multinormality (von Eye & Gardiner, 2004).

Many data analysts tend to believe that such methods as MANOVA are robust against violations of multinormality or against heteroscedasticity. Indeed, Ito (1980) states that rather strong evidence exists that the ANOVA F-test which is derived under the assumptions of normality and homoscedasticity is „extremely robust under violations of these assumptions" (p. 220). This applies accordingly to the tests used for MANOVA. However, Ito states also that the „major exception to this statement occurs for small and unequal sample sizes" (p. 220). We note that this is the typical case of applications of such methods in the empirical social sciences. The major exception is thus the norm, which makes it even more important to be able to diagnose multinormality violations.

1. A brief review of tests of multinormality

The behavior of tests under violations of normality assumptions is a classic topic in the theoretical and the applied statistical literature. Early work was performed by Pearson (1931). Current textbooks either cite the results that suggest robustness, just mention the assumption of multinormality (Bartholomew, Steele, Moustaki, & Galbright, 2002), or describe a selection of tests of multinormality (Jobson, 1992). Two of the more frequently employed methods are briefly reviewed in this section. The first method plots individual data points against points that are expected under the assumption of multinormality. Specifically, consider the squared Mahalanobis distance of data vector x_i from its sample mean, $\bar{x}$,

$$m_i^2 = (x_i - \bar{x})' S^{-1} (x_i - \bar{x}),$$

where x_i is the data vector of case i, and S is the sample covariance matrix. The N multivariate distances m_i^2 can be ordered and plotted against the χ^2 distribution percentiles

$\chi^2_{(1-\alpha);d}$ with $(1 - \alpha_i) = (i - 0.5)/N$, and $i = 1, 2, ..., N$. The scatterplot of the points $[m_i^2, \chi_{(1-\alpha_i);d}]$ should show a straight line.

Most popular, and available in a number of the general purpose statistical software packages, e.g., in S+, are Mardia's (1970, 1980) measures of multivariate skewness and kurtosis. The sample measure of multivariate skewness is

$$\hat{\gamma}_{1d} = \frac{1}{N^2}\sum_{i,j} m_{ij}^3,$$

where $m_{ij} = (x_i - \bar{x})' S^{-1} (x_j - \bar{x})$. The statistic $N\hat{\gamma}_{1d}/6$ follows a χ^2 distribution with df = $d(d + 1)(d + 2)/6$.

Mardia's multivariate kurtosis is

$$\hat{\gamma}_{2d} = \frac{1}{N}\sum_{i,j} m_i^4,$$

where $m_i^2 = (x_i - \bar{x}) S^{-1} (x_i - \bar{x})$, and $i = 1, ..., N$. The statistic $\hat{\gamma}_{2d}$ is normally distributed with mean $d(d + 2)$ and variance $8d(d + 2)/N$, where d is the number of variables under study. An early, univariate precursor of this test was proposed by David, Hartley, and Pearson (1954).

These two measures allow one to test two hypotheses that are compatible with the assumption of multinormality. Specifically, if a sample was randomly drawn from a multinormal population, there should be no significant skew, and the measure of kurtosis should deviate from expectancy, that is from 0, only randomly[2]. If these measures suggest significant deviations, one can conclude with reasonable certainty that the data come from a non-normal population (other implications are discussed below). However, it is unknown whether particular variable patterns or sectors in the multivariate space exist that carry an inordinate amount of the deviation. In the following section, χ^2-tests are presented that allow one to test hypotheses about particular deviations.

2. χ^2-tests of deviations from normality

In this section, we first illustrate the well-known χ^2-test for univariate normality. When only one variable is studied, this test is most useful. When multiple variables are studied, multinormality tests are typically preferred. The reason for this choice is that individual variables from a set of variables that are jointly multinormally distributed, are also normally distributed. In contrast, if a number of variables is normally distributed individually, they are not necessarily also multinormal, unless they are independent of each other. Second, we present the new sector test of multinormality and its companion omnibus test.

[2] It should be noted, however, that even if skewness and kurtosis are 0, one cannot be sure that normality exists. There exist (1) an extremely non-normal three point distribution at (0, 0), and (2) infinitely many other distributions that are non-normal.

2.1 χ^2 test of univariate normality

The χ^2 goodness-of-fit test of univariate normality is well known and does not need to be described in detail. The test involves the following steps:

1) segmenting the variable under study;
2) counting the number of cases per segment;
3) calculating z-scores for the boundaries of the segments;
4) estimating the probability for each segment (for more detail, see Point 3 in Section 2.2, below);
5) calculating the estimated expected segment frequencies;
6) comparing the observed and the estimated expected segment frequencies using the Pearson goodness-of-fit χ^2-test.

In the following paragraphs, we illustrate this test using a data set that was collected as part of a larger project on parenting (Bogat et al., 1998). The participants in this project were 175 pregnant adolescents in an alternative school in Lansing, Michigan. The average age of the adolescents was about 15 years (range: 12 - 20). Their ethnical backgrounds were 29% Caucasian, 43% African American, 14% Latina, 3% Native American, 1% Asian American, and 10% other/multiracial. The school served predominantly low income families (approximately 60% received public assistance). Most referrals to the program were made by school counselors primarily because they deemed the adolescents at-risk of dropping out of school (e.g., poor grades, sporadic school attendance). The study itself had several goals including assessing the influence of mental health (e.g., anxiety and depression), developmental history (e.g., autonomy from parents), academic achievement, and social support on the adolescent's parenting abilities as well as her confidence to cope with the vicissitudes of parenting. The data presented here and in the second example, below, were collected while the adolescent was pregnant and within 10 days of entering the alternative school.

One element of this project involved testing hypotheses on the relationship between mental health and parenting. One of the indicators of mental health was T1GHQ. To be suitable for multivariate data analysis, this variable has to be normally distributed by itself. To examine this variable in regard to its distributional characteristics, we first present two graphical representations and then perform the χ^2 goodness-of-fit test. The graphical representations appear in Figure 1.

The bar chart of T1GHQ in Figure 1 shows the bars and a normal distribution blended in. The normal distribution indicates how the variable would be distributed were its distribution normal. Clearly, there is a skew. This skew is significantly different than zero: skewness = 0.582; ASE = 0.184; z = 3.16; p < 0.01. This result is presented in a slightly different way in the second panel of Figure 1: the observed distribution appears on the top border of the box, the expected distribution appears on the right hand border, and the scatterplot appears inside the box. The line deviates from the straight line that describes a normally distributed variable.

The following paragraphs present the χ^2 analysis to answer the question whether variable T1GHQ is normally distributed. The variable is scaled to have a range from 1.5 through 4.0 with a mean of 2.48, and sd = 0.526. Five segments were created, the first ending at 2.00 (p = 0.149), the second ending at 2.50 (p = 0.334), the third ending at 3.00 (p = 0.323), the fourth

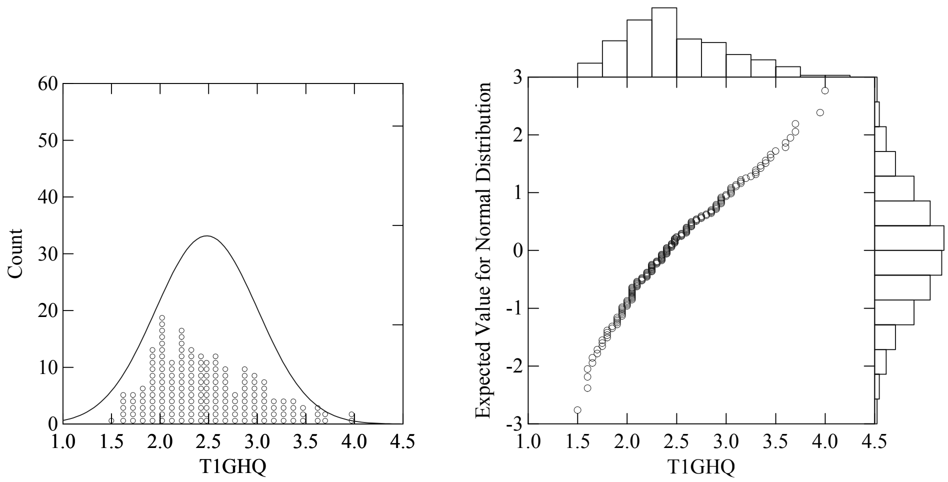

Figure 1:
Bar chart (left panel) and q-plot (right panel) of the variable T1GHQ

ending at 3.50 (p = 0.135), and the last ending at 4.0 (p = 0.024). These segments are equidistant on the raw score scale. The probabilities reflect again that the variable is skewed, because otherwise the segments would mirror at the mean and would be equiprobable. Table 1 displays the χ^2 analysis.

Table 1:
χ^2 analysis to test the hypothesis that variable T1GHQ is normally distributed

Segment	m_i	$\hat{m}_i$	X^2	p
1	34	26.145	2.360	.1245
2	70	58.485	2.267	.1321
3	42	56.595	3.764	.0524
4	22	23.695	.121	.7277
5	7	4.27	1.745	.1865
sums	175	175	10.257	

The overall Pearson X^2 has 3 degrees of freedom and comes with p = 0.0165. This small probability suggests that variable T1GHQ is not normally distributed. For the distributional characteristics of this variable, see Figure 1. In the next section, we describe the new sector test for multinormality.

2.2 A sector test of multinormality

The following sector test can be seen as a multivariate extension of the univariate χ^2 test that was illustrated in the last section. Specifically, the multivariate version of this test proceeds in the following steps:

1) *Segment each of the d variables under study*. For the jth variable, we obtain c_j segments, with $j = 1, ..., d$.
2) *Crossing the segmented variables*. Crossing all segmented variables yields a d-dimensional cross-classification with $\prod_{j=1}^{d} c_j$ sectors. In the next step, the probability needs to be calculated for a case to be located in one of these d-dimensional sectors.
3) *Calculating the probability of each sector*. Consider the univariate case first. Let the first segment to have boundaries $-\infty$ and z_1. Then, the probability for an element to be located in this segment is

$$p(z_{-\infty}) - p(z_1) = 0 - \int_{-\infty}^{z_1} \Psi(z)dz,$$

that is, the area under the normal curve from $-\infty$ to z_1. In more general terms, let the boundaries of a segment be z_k and z_{k+1}. The area under the normal curve for this segment is

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

Now, let p_k^j denote the probability that an element sits in segment k of the jth variable. In the multivariate case, cases sit in the sectors that were created by crossing the 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_k^d and z_{k+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

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

Until 1992, there was no easily tractable solution for this equation. Genz (1992) proposed such a solution (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). Somerville (1998) proposed a solution for convex, that is, ellipsoid sectors. For the test that is described in this article, we use Genz' solution. We label the individual sector $s_{i, j,...,k}$, and abbreviate the probability of sitting in a d-dimensional right-angle sector with $p_{i, j,...,k}$.

4) *Estimating expected sector frequencies.* The expected frequency of objects in Sector $s_{i,j,...,k}$ is $e_{i,j,...,k} = Np_{i,j,...,k}$. The next step involves performing tests for each individual sector.
5) *Sector-specific tests*. To identify locations of violations of multinormality, one compares for each Sector $s_{i,j,...,k}$ the observed frequency of objects, $o_{i,j,...,k}$ with the corresponding expected frequency, $e_{i,j,...,k}$ under the null hypothesis $E[o_{i,j,...,k}] = e_{i,j,...,k}$. If this comparison suggests that a sector contains significantly more or fewer objects than expected based on the joint density function of the d variables under study, this sector evinces a violation of multivariate normality. Therefore, the assumption of multivariate normality must be rejected at least for this sector. There is a large number of tests that are suitable for the present purpose (von Eye, 2002). Here, we select the well-known Pearson X^2-component test, for three reasons. First, the X^2-components are known to have desirable properties for the analysis of individual cells of a multivariate cross-classification (von Eye, 2002; von Weber, Lautsch, & von Eye, 2004). Second, the component test is directly parallel to the univariate χ^2-test. Third, the components sum up to an omnibus test statistic. The test statistic for the individual sector is

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

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 $\alpha^* = \alpha / \prod_{j=1}^{d} c_j$. This threshold becomes rapidly impractical as the number of sectors increases. Therefore, more efficient procedures such as the ones proposed by Holm (1979) or Keselman, Cribbie, and Holland (1999) may be preferable, because they yield protected α-thresholds that are less prohibitive.
6) *Performing an omnibus test*. The sum of the X^2-components yields the omnibus test statistic

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

This statistic can be used as a test of whether, overall, the cross-classification of segments follows a multinormal distribution. The test has

$$df = \left(\prod_{j=1}^{d} c_j \right) - 2d - d_{cov} - 1 ,$$

where c_j is the number of segments of the jth variable. The term d_{cov} indicates the number of correlations (or covariances) taken into account. Typically, $d_{cov} = \binom{d}{2}$, that is, all covariances are taken into account.

Determining the spacing of segments. The spacing of segments is an issue of concern in applications of the present test of multinormality. Two concepts of spacing have been discussed (von Eye & Gardiner, 2004), *equidistant spacing* and *equiprobability spacing*.

Equidistant spacing involves creating segments that span the same distance on a raw score scale. In other words, equidistant spacing splits the distance between the two extreme scores in equal segments. The advantage of this procedure is that the segments have a natural interpretation. However, there is a big drawback. The extreme segments come with small probabilities. Consider a study in which the three variables neuroticism, schizophrenia, and use of leisure drugs are studied. Now suppose that the lowest segments of each of these three scales have an upper bound of T = 20, that is, separate the bottom 13% of the population from the top 87%. The joint probability of these three segments is $p = 0.0022$ (possible correlations among the three variables not taken into account). This applies accordingly to the segments close to the upper end of the scale. The sample needed for the X^2-tests to perform well is so large that the test may become inapplicable.

Therefore, equiprobability spacing has been proposed. Here, the probability scale is segmented instead of the raw score scale. Splitting the probability scale in equally-spaced segments comes with the advantage over equidistant segmenting that the segments have equal a priori-determined probabilities. Thus, the segments at the extremes of the scales have the same probability as the segments close to the midpoints of the scales. This characteristic simplifies calculation. On the downside is that the segments will differ in length on the raw score scale. Specifically, the segments close to the midpoint of the raw score scale will be short, whereas the segments towards the ends of the raw score scales will be long. Examples follow below.

Determining the number of segments. When determining the number of sectors, two arguments are of importance, sample size and statistical power. When the sample size is given, one can use power calculations to determine the number of sectors. For a given effect size, one can calculate the number of cases needed for each test, and then determine the number of sectors. For example, to be able to detect small deviations from multinormality ($w = 0.1$) using a χ^2-test , one needs 1000 cases per sector for $p = 0.8$ (Cohen, 1988). If one trusts in the robustness of statistical methods and looks for large effects only ($w = 0.6$), the number of 30 cases per sector may be sufficient for $p = 0.8$. Here again, large numbers of segments, and segments that are equidistant on the raw score scale may be problematic because of possibly burdensome sample size requirements.

3. Data examples

For the following data examples, we use the same data as for the first example, above. One question that the project asked concerned the relationship among indicators of autonomy and coping. Specifically, it was asked whether autonomy information allows one to predict the ability to cope. For the following illustration, we examine the two autonomy indicators, T1ATTAUT (attitudinal autonomy) and T1EMOAUT (emotional autonomy), and the summary scale of coping, T1COPALL. We proceed in three steps. First, we calculate descriptive information for each of these three variables. Second, we perform uni- and three-variate graphical analyses to gain visual insight into the uni- and three-variate distributions of the three variables, and third, we ask whether the three-dimensional distribution of the three variables is multinormal. The descriptive statistics for the three variables appear in Table 2.

Table 2:
Descriptive statistics for three variables of autonomy and coping

	T1EMOAUT	T1ATTAUT	T1COPALL
N of cases	175	175	175
Minimum	1.071	1.125	1.273
Maximum	5.000	5.000	3.182
Mean	3.260	3.037	2.244
Standard Dev	0.919	0.815	0.400
Skewness(G1)	-0.264	-0.017	-0.110
SE Skewness	0.184	0.184	0.184
Kurtosis(G2)	-0.568	-0.520	-0.382
SE Kurtosis	0.365	0.365	0.365

The statistics in Table 2 suggest that the three variables may follow a univariate normal distribution. Specifically, their skewness and kurtosis scores do not differ significantly from expectancy. This conclusion is supported by the graphs displayed in Figure 2.

Figure 2 suggests that the three variables are not perfectly well behaved. However, deviations from the univariate normal distribution are certainly not dramatic and we conclude, again, that these variables are normally distributed. We now ask whether this conclusion holds when we inspect the 3D scatterplot, in Figure 3.

The visual inspection of Figure 3 suggests no obvious deviations from the ellipsoid form of the data cloud that one would expect from three multinormal, moderately to weakly correlated variables ($r_{T1EMOAUT-T1ATTAUT} = 0.48$; $r_{T1EMOAUT-T1COPALL} = 0.09$; $r_{T1ATTAUT-T1COPALL} = 0.09$). Thus, to find out whether these three variables deviate significantly from a three-variate normal distribution, we have to use statistical tests. We use four tests, Mardia's test of multivariate kurtosis, Mardia's test of multivariate skewness, the sector tests proposed here, and the χ^2 omnibus test also proposed in this article. In addition, we perform the sector and omnibus tests using the two methods of segmenting discussed in the last section, that is, segmenting that is equidistant on the raw score scale, and segmenting that is equidistant on the probability scale. In both cases, we take the correlations among the three variables into account, to

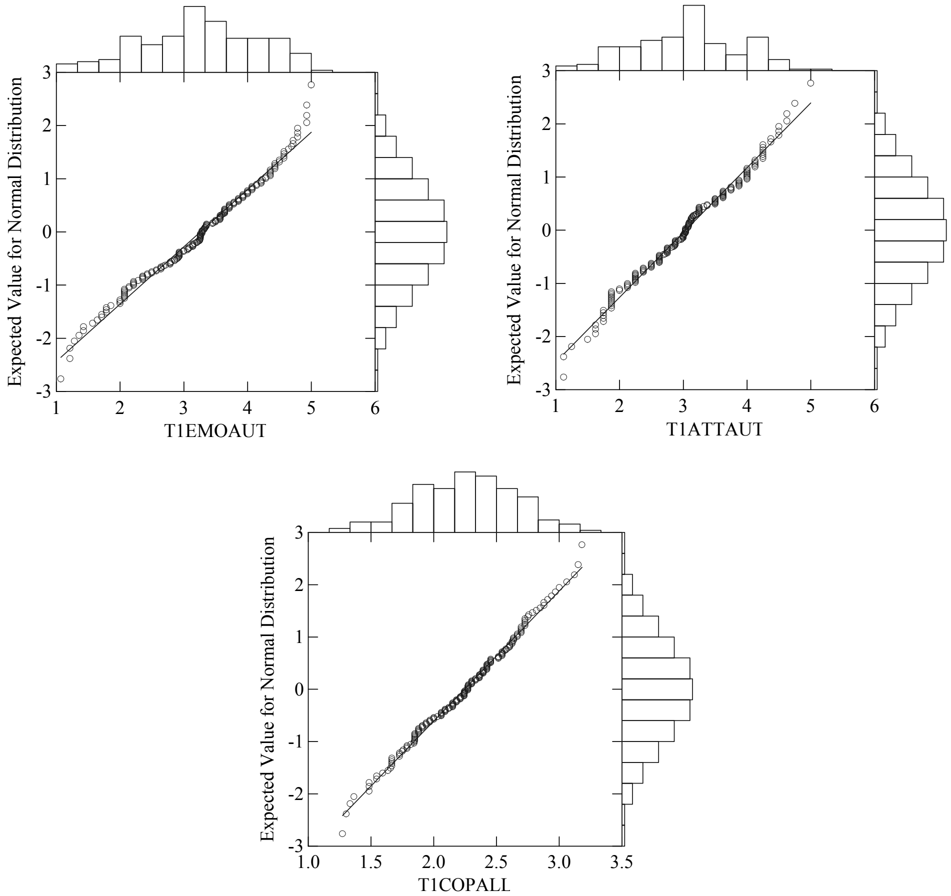

Figure 2:
q-plots of two autonomy and one coping variable

make sure significant deviations indicate sectors of non-normality rather than variable correlations (as would be of interest in applications of Configural Frequency Analysis; von Eye, 2002).

Mardia's tests. For the three variables, T1EMOAUT, T1ATTAUT, and T1COPALL, we calculate the value of 0.184 for the multivariate skewness (p = 0.865) and the value of 12.834 for the multivariate skewness (z = -2.616; p = 0.0045). We thus conclude that whereas there is no significant skew, the three-variate distribution is slightly heavy around the belt line (deCarlo, 1997). The assumption of multinormality can thus be rejected. We now ask whether individual sectors exist that carry an inordinate portion of this deviation from multinormality.

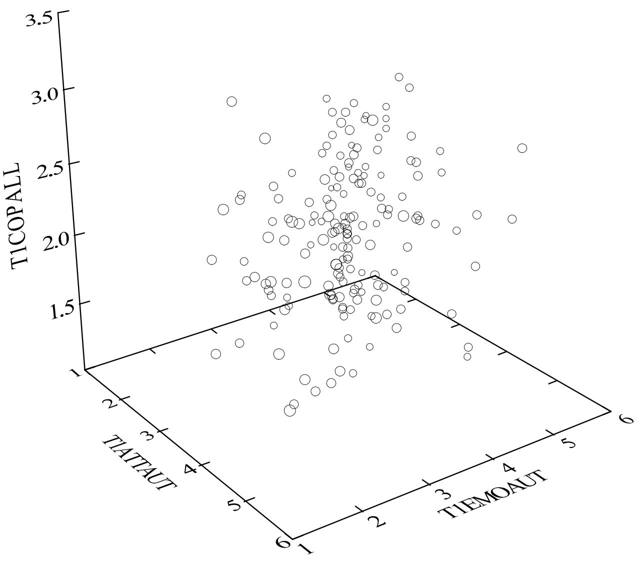

Figure 3:
3D scatterplot of two autonomy and one coping variable

Testing multinormality based on segments that are equidistant on the raw score scale. For both runs of the sector test, we created three segments on each of the variables. This led to 27 sectors in the three-dimensional data space and an average of 175/27 = 6.48 cases per sector (under a null model). The power of each individual sector test will thus be small. Still, there may be strong deviations. Table 3 displays the boundaries of the segments for each variable and the segment probabilities. The boundaries at the lower ends of the scales are not repeated (see minima in Table 2). The Pearson X^2-component test was used to test each sector hypothesis, and the significance level was Bonferroni-adjusted to be $\alpha^* = .05/27 = 0.00185$.

Table 4 displays the results of the sector tests.

Table 3:
Segment boundaries and segment probability estimates for segments that are equidistant on the raw score scales

Variable	Segment boundaries			Segment Probabilities		
T1EMOAUT	2.38	3.69	5.00	.161	.511	.291
T1ATTAUT	2.42	3.71	5.00	.214	.572	.197
T1COPALL	1.91	2.55	3.18	.194	.573	.216

Table 4:
Sector X^2-tests based on segments that are equidistant on the raw score scale, correlations taken into account

Sector Index	Frequencies Observed	Expected	X^2-statistic	p
111	6	1.975	8.204	.00417976
112	7	4.394	1.545	.21380364
113	8	2.271	14.451	.00014382*
121	3	4.373	.431	.51148416
122	7	9.730	.766	.38151161
123	4	5.029	.210	.64638712
131	0	1.881	1.881	.17024146
132	1	4.185	2.424	.11950667
133	0	2.163	2.163	.14137414
211	3	4.443	.469	.49349759
212	8	9.887	.360	.54849609
213	2	5.110	1.893	.16889434
221	11	9.839	.137	.71128922
222	33	21.892	5.636	.01759101
223	10	11.315	.153	.69587554
231	4	4.232	.013	.91026967
232	6	9.416	1.239	.26562870
233	4	4.867	.154	.69444052
311	3	3.182	.010	.91885760
312	4	7.079	1.339	.24713630
313	1	3.659	1.932	.16451013
321	5	7.045	.594	.44098138
322	12	15.676	.862	.35321836
323	8	8.102	.001	.97141188
331	5	3.030	1.280	.25781107
332	11	6.742	2.689	.10105231
333	9	3.485	8.729	.00313193

The results in Table 4 suggest that Sector 113 contains more cases than compatible with the assumption of a three-variate normal distribution. These are individuals with low scores on both attitudinal variables and a high score on coping. Two additional sectors may carry more cases than expected (Sectors 111 and 333), but the conservative Bonferroni adjustment in combination with the low power for these two tests may have prevented these discrepancies from becoming significant.

The omnibus X^2-score is 59.567 ($df = 27 - 6 - 3 - 1 = 17$; $p < 0.01$); this suggests that the expected distribution differs significantly from the observed one.

Testing multinormality based on segments that are equidistant on the probability scale. We now present the results that are based on equiprobable segments. The testing itself used the same procedures as for the run with equidistant segments. Table 5 displays the results from the sector tests.

The results in Table 5 differ from the ones in Table 4 in two important respects. First, the estimated expected cell frequencies are closer to a uniform distribution. In fact, these frequencies are uniform as follows from the specification of the equiprobable segments on the probability scale. For correlations equal to zero, each of the expected frequencies would be $e_{ijk} = 175/27 = 6.48$. The deviations from this value that are obvious in Table 5, reflect the variable intercorrelations. Second, there is not a single sector that evinces a significant deviation from multinormality. All together, however, the deviations amount to a significant discrepancy ($X^2 = 55.271$; $df = 19$; $p < 0.01$).

From these results, we conclude that

1) the sector tests, in tandem with the X^2 omnibus test, are a powerful tool for the detection of deviations from multinormality;
2) the sector tests allow one to search for sectors in the multivariate space that carry an inordinate portion of the overall deviation from multinormality;
3) even if no single sector stands out as indicating particularly blatant violations of multinormality, the omnibus test can still suggest that, overall, the assumption of multinormality must be rejected;
4) the selection of segments may influence the appraisal of the multinormality hypotheses.

4. Discussion

Multinormality is a condition that often must be met to obtain parameter estimates that are efficient and unbiased. In many social science applications, robustness cannot be assumed because sample sizes tend to be small. Therefore, it is recommended to perform tests of multinormality before applying parametric multivariate statistical procedures. In this article, we discuss known tests of multinormality, specifically, Mardia's tests of multivariate skewness and multivariate kurtosis. These tests allow one to test hypotheses that conform with the assumption of multinormality. However, these tests do not allow one to test hypotheses about multinormal density directly. In addition, these tests provide no information about whether violations are prominent specifically in particular sectors of the multivariate space. The newly proposed sector tests and the companion omnibus test do allow one to test hypotheses that address these issues. What are the implications of the results of such tests?

Table 5:
Sector X^2-tests based on segments that are equidistant on the probability scale, correlations taken into account

Sector Index	Frequencies Observed	Expected	X^2-statistic	p
111	9	5.409	2.384	.12256823
112	9	5.129	2.921	.08741991
113	12	5.782	6.687	.00971149
121	3	6.182	1.638	.20066252
122	6	5.862	.003	.95450639
123	1	6.608	4.759	.02914093
131	4	5.312	.324	.56910070
132	3	5.038	.824	.36397446
133	4	5.679	.496	.48115553
211	7	7.000	.000	.99993107
212	1	6.638	4.788	.02865255
213	4	7.483	1.621	.20297573
221	16	8.000	8.001	.00467571
222	12	7.586	2.568	.10901820
223	12	8.551	1.391	.23828623
231	3	6.875	2.184	.13945990
232	4	6.519	.973	.32381419
233	7	7.349	.017	.89759298
311	5	6.151	.215	.64250081
312	5	5.833	.119	.73012480
313	4	6.576	1.009	.31519005
321	3	7.030	2.310	.12852033
322	6	6.666	.067	.79631524
323	5	7.515	.842	.35893172
331	8	6.041	.635	.42555684
332	9	5.729	1.868	.17174763
333	13	6.458	6.627	.01004601

The implications are relatively clear when null hypotheses concerning multinormality are rejected. In this case, one can assume that the data at hand have a very small probability to have been drawn from a multinormal population. If the sector test identifies one or several sectors as carrying inordinate amounts of the deviations from multinormality, one can ask whether there may be a reason for this imbalance. Possible reasons include sampling problems, selective sampling, and mis-specification of populations. Researchers may then consider remedial steps, for instance, completing a sample by searching for cases under-represented in the first data collection. In the data example given in the previous section, one may ask why too many cases with low scores on both attitudinal variables and a high score on coping are members of the sample (Pattern 113; Table 4). If inadvertent selective procedures led to oversampling of such cases, resampling may be considered.

This last issue highlights that the proposed tests are useful from two perspectives. First, they allow one to test multinormality assumptions. Second, they allow one to determine where in the sample there may be under- or overrepresented groups of cases. Thus, even if one assumes that certain tests are robust, the tests proposed here, in particular the sector test, provide the means of identifying those sectors of the data space that may need to be resampled to obtain a representative sample from a multinormal population.

Interesting is the fact that in those cases in which deviations from multinormality are most deleterious, the tests proposed here (as well as a number of other tests of multinormality (see Mardia, 1980)) may have the least power. These are the cases in which sample sizes are small. In other words, when needed the most, the existing tests may perform the worst. There are several ways out of this problem. First, the sample can be increased. Second, one can use more powerful tests. Examples of such tests have been discussed in the context of Configural Frequency Analysis (von Eye, 2002; see also von Eye & Mun, 2003; von Weber, von Eye, & Lautsch, 2004). A most powerful test was proposed by Lehmacher (1981). This test, however, can be applied only when variable correlations are not taken into account, a rather unrealistic situation for the present purposes. A third way involves employing exact tests. Finally, if certain sectors are suspected to carry the brunt of non-normality, testing can focus on these sectors. As a consequence, the number of tests becomes smaller, and the protected α-level becomes less extreme.

If, based on the proposed tests, researchers conclude that patterns with too many or too few cases indicate that the concept of multinormality may not apply, alternative statistical methods need to be found. Alternatives include robust methods, methods that are based on more flexible distributional assumptions, and weighted least squares methods. Unfortunately, alternative methods are not always available.

In the case in which none of the tests leads one to reject null hypotheses concerning multinormality, researchers need to be cautious nevertheless. In this case, the results of the tests only suggest that the null hypotheses concerning multinormality remain un-rejected. These results provide no proof that multinormality exists in the population. The same sampling errors that can result in selective samples that contradict multinormality although it exists in the population, can lead to collecting data that suggest multinormality because the sample is biased. If this is the case, tests of multinormality may be useless. The same applies to the parameters estimated under such conditions. Still, if multinormality hypotheses can be retained, parametric statistical procedures are typically applied without much hesitation.

References

1. Bartholomew, D.J., Steele, F., Moustaki, I., & Galbraith, J.I. (2002): The analysis and interpretation of multivariate data for social scientists. Boca Raton: Chapman & Hall.
2. Bogat, G.A., Caldwell, R.A., Guzmán, B., Galasso, L., & Davidson, W.S.II (1998): Structure and stability of maternal support among pregnant and parenting adolescents. Journal of Community Psychology, 26, 549-568.
3. Cohen, J. (1988): Statistical power analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Erlbaum.
4. David, H.A., Hartley, H.O., & Pearson, E.S. (1954): The distribution of the ratio, in a single normal sample, of range to standard deviation. Biometrika, 41, 482-493.
5. DeCarlo, L.T. (1997): On the meaning and use of kurtosis. Psychological Methods, 2, 292-307.
6. Genz, A. (1992): Numerical computation of multivariate normal probabilities. Journal of Computational and Graphical Statistics, 1, 141-149.
7. Gupta, S.S. (1963): Probability integrals of multivariate normal and multivariate t. The Annals of Mathematical Statistics, 34, 792-828.
8. Holm, S. (1979): A simple sequentially rejective multiple test procedures based on a modified Bonferroni test. Scandinavian Journal of Statistics, 6, 65-70.
9. Ito, P.K. (1980): Robustness of ANOVA and MANOVA test procedures. In P.R. Krishnaiah (ed.), Handbook of statistics 1. Analysis of variance (pp. 199-236). Amsterdam: North Holland.
10. Jobson, J.D. (1992): Applied multivariate data analysis. Volume II: Categorical and multivariate analysis. New York: Springer.
11. 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.
12. Lehmacher, W. (1981): A more powerful simultaneous test procedure in Configural Frequency Analysis. Biometrical Journal, 23, 429-436.
13. Mardia, K.V. (1970): Measures of multivariate skewness and kurtosis with applications. Biometrika, 57, 519-530.
14. Mardia, K.V. (1980): Tests of univariate and multivariate normality. In P.R. Krishnaiah (ed.), Handbook of statistics (vol. 1; pp 279-320). Amsterdam: North Holland.
15. 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
16. Pearson, E.S. (1931): The analysis of variance in cases of non-normal variation. Biometrika, 23, 114-133.
17. 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.
18. 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.
19. von Eye, A. (2002): Configural Frequency Analysis: methods, models, and applications. Mahwah, NJ: Lawrence Erlbaum.
20. von Eye, A., & Gardiner, J.C. (2004): Locating deviations from multivariate normality. Understanding statistics (in press).
21. von Eye, A., & Mun, E.-Y. (2003): Characteristics of measures for 2 x 2 tables. Understanding Statistics, 2, 243-266.
22. 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. (In press)