

# On Mardia's Tests of Multinormality

Kankainen, A., Taskinen, S., and Oja, H.

**Abstract.** Classical multivariate analysis is based on the assumption that the data come from a multivariate normal distribution. The tests of multinormality have therefore received very much attention. Several tests for assessing multinormality, among them Mardia's popular multivariate skewness and kurtosis statistics, are based on standardized third and fourth moments. In Mardia's construction of the affine invariant test statistics, the data vectors are first standardized using the sample mean vector and the sample covariance matrix. In this paper we investigate whether, in the test construction, it is advantageous to replace the regular sample mean vector and sample covariance matrix by their affine equivariant robust competitors. Limiting distributions of the standardized third and fourth moments and the resulting test statistics are derived under the null hypothesis and are shown to be strongly dependent on the choice of the location vector and scatter matrix estimate. Finally, the effects of the modification on the limiting and finite-sample efficiencies are illustrated by simple examples in the case of testing for the bivariate normality. In the cases studied, the modifications seem to increase the power of the tests.

## 1. Introduction

Classical methods of multivariate analysis are for the most part based on the assumption that the data are coming from a multivariate normal population. In the one-sample multinormal case, the regular sample mean vector and sample covariance matrix are complete and sufficient, but on the other hand highly sensitive to outlying observations. It is also well known that the tests and estimates based on the sample mean vector and sample covariance matrix have poor efficiency properties in the case of heavy tailed noise distributions. Testing for departures from multinormality helps to make practical choices between competing methods of analysis.

In the univariate case, standardized third and fourth moments  $b_1$  and  $b_2$  are often used to indicate the skewness and kurtosis. For a random sample  $\mathbf{x}_1, \dots, \mathbf{x}_n$  from a  $p$ -variate distribution with sample mean vector  $\bar{\mathbf{x}}$  and sample covariance

---

Received by the editors May 2, 2003.

1991 *Mathematics Subject Classification.* Primary 62H15; Secondary 62F02.

*Key words and phrases.* Kurtosis, Multinormality, Robustness, Skewness.

matrix  $S$ , Mardia (1970,1974,1980) defined the  $p$ -variate skewness and kurtosis statistics as

$$b_{1,p} = \text{ave}_{i,j}[(\mathbf{x}_i - \bar{\mathbf{x}})^T S^{-1}(\mathbf{x}_j - \bar{\mathbf{x}})]^3$$

and

$$b_{2,p} = \text{ave}_i[(\mathbf{x}_i - \bar{\mathbf{x}})^T S^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})]^2,$$

respectively. The statistics  $b_{1,p}$  and  $b_{2,p}$  are functions of standardized third and fourth moments, respectively. From the definition one easily sees that  $b_{1,p}$  and  $b_{2,p}$  are invariant under affine transformations. In the univariate case these reduce to the usual univariate skewness and kurtosis statistics  $b_1$  and  $b_2$ . Mardia advocated using the skewness and kurtosis statistics to test for multinormality as they are distribution-free under multinormality. See also Bera and John (1983) and Koziol (1993) for the use of the standardized third and fourth moments in the test construction.

Intuitively, as in the univariate case, the skewness and kurtosis statistics  $b_{1,p}$  and  $b_{2,p}$  compare the variation measured by third and fourth moments to that measured by second moments. The second, third and fourth moments are all highly non-robust statistics, however, and more efficient procedures may be obtained through comparisons of robust and nonrobust measures of variations. It is therefore an appealing idea to study whether it is useful or reasonable to replace the regular sample mean vector  $\bar{\mathbf{x}}$  and sample covariance matrix  $S$  by some affine equivariant robust competitors, location vector estimate  $\hat{\boldsymbol{\mu}}$  and scatter matrix estimate  $\hat{\Sigma}$ .

The paper is organised as follows. In Section 2 the limiting distributions of Mardia's test statistics under the null model of multinormality are discussed. The test statistics are based on the standardized third and fourth moments; the regular mean vector and regular covariance matrix are used in the standardization. Section 3 lists the tools and assumptions for alternative  $\sqrt{n}$ -consistent affine equivariant location and scatter estimates. The location and scatter estimates are then used in the regular manner to standardize the data vectors. The limiting distributions of the standardized third and fourth moments in this general case are found in Section 4. As an illustration of the theory, the effect of the way of standardization on the limiting and small sample efficiencies in the bivariate case are studied in Sections 5 and 6. The auxiliary lemmas and proofs have been postponed to the Appendix.

## 2. Limiting null distributions of Mardia's statistics

In this section we recall the wellknown results concerning the limiting null distributions of the Mardia's skewness and kurtosis statistics. As, due to affine invariance, the distributions of the test statistics do not depend on the unknown  $\boldsymbol{\mu}$  and  $\Sigma$ , it is not a restriction to assume in the following derivations that  $\mathbf{x}_1, \dots, \mathbf{x}_n$  is a random sample from the  $N_p(\mathbf{0}, I_p)$  distribution. Write

$$\mathbf{z}_i = S^{-1/2}(\mathbf{x}_i - \bar{\mathbf{x}}), \quad i = 1, \dots, n,$$

for the data vectors standardized in the classical way. Mardia's skewness statistic can be decomposed as

$$b_{1,p} = 6 \sum_{j < k < l} [\text{ave}_i \{z_{ij} z_{ik} z_{il}\}]^2 + 3 \sum_{j \neq k} [\text{ave}_i \{z_{ij}^2 z_{ik}\}]^2 + \sum_j [\text{ave}_i \{z_{ij}^3\}]^2$$

and the Mardia's kurtosis statistic is similarly

$$b_{2,p} = \sum_{j \neq k} \text{ave}_i \{z_{ij}^2 z_{ik}^2\} + \sum_j \text{ave}_i \{z_{ij}^4\}.$$

See e.g. Koziol (1993). Under multinormality  $b_{1,p}$  and  $b_{2,p}$  are affine invariant, all the standardized third and fourth moments are asymptotically normal and asymptotically independent and consequently the limiting distributions of

$$n \frac{b_{1,p}}{6} \quad \text{and} \quad \sqrt{n} \frac{b_{2,p} - p(p+2)}{\sqrt{8p(p+2)}}$$

are a chi-square distribution with  $p(p+1)(p+2)/6$  degrees of freedom and a  $N(0, 1)$  distribution, respectively.

### 3. Location and scatter estimates

In Mardia's test construction, the data vectors are standardized using the regular sample mean vector  $\bar{\mathbf{x}}$  and sample covariance matrix  $S$ . In this paper we thus consider what happens if these are replaced by other  $\sqrt{n}$ -consistent affine equivariant location and scatter estimates of  $\boldsymbol{\mu}$  and  $\Sigma$ , say,  $\hat{\boldsymbol{\mu}}$  and  $\hat{\Sigma}$ . Next we list assumptions and useful notations for location vector estimate  $\hat{\boldsymbol{\mu}}$  and scatter matrix estimate  $\hat{\Sigma}$ .

First we assume that, for the  $N_p(\mathbf{0}, I_p)$  distribution, the influence functions of these location and scatter functionals at  $\mathbf{z}$  are

$$\gamma(r) \mathbf{u} \quad \text{and} \quad \alpha(r) \mathbf{u} \mathbf{u}^T - \beta(r) I_p,$$

respectively, where  $r = \|\mathbf{z}\|$  and  $\mathbf{u} = \|\mathbf{z}\|^{-1} \mathbf{z}$ . We will later see that the limiting distributions of the modified test statistics depend on the used location and scatter estimates through these functions. For the influence functions of location and scatter functionals, see also Hampel et al. (1986), Croux and Haesbroek (2000) and Ollila et al. (2003a,b).

We further assume that, again under the  $N_p(\mathbf{0}, I_p)$  distribution,

$$\mathbf{t} = \sqrt{n} \text{ave}\{\gamma(r_i) \mathbf{u}_i\} = \sqrt{n} \hat{\boldsymbol{\mu}} + o_P(1)$$

and

$$\mathbf{C} = \sqrt{n} \text{ave}\{\alpha(r_i) \mathbf{u}_i \mathbf{u}_i^T - \beta(r_i) I_p\} = \sqrt{n} (\hat{\Sigma} - I_p) + o_P(1).$$

Finally, assume that, for the  $N_p(\mathbf{0}, I_p)$  distribution, the expected values  $E(\gamma^2(r))$ ,  $E(\alpha^2(r))$  and  $E(\beta^2(r))$  are finite. These assumptions imply that the limiting distributions of  $\sqrt{n} \hat{\boldsymbol{\mu}}$  and  $\sqrt{n} \text{vec}(\hat{\Sigma} - I_p)$  are  $p$ - and  $p^2$ -variate normal distributions

with zero mean vectors and covariance matrices

$$\frac{E[\gamma^2(r)]}{p} I_p$$

and

$$\frac{E[\alpha^2(r)]}{p(p+2)} [I_{p^2} + I_{p,p} + \text{vec}(I_p) \text{vec}(I_p)^T] + \left( E[\beta^2(r)] + \frac{2}{p} E[\alpha(r)\beta(r)] \right) \text{vec}(I_p) \text{vec}(I_p)^T,$$

respectively. (Here  $I_{p,p}$  is the so called commutation matrix. See e.g. Ollila et al. (2003b).)

#### 4. Limiting distribution of the standardized third and fourth moments

Consider now any estimators  $\hat{\boldsymbol{\mu}}$  and  $\hat{\Sigma}$  and write

$$\mathbf{z}_i = \hat{\Sigma}^{-1/2}(\mathbf{x}_i - \hat{\boldsymbol{\mu}}), \quad i = 1, \dots, n,$$

for the standardized observations. Denote the standardized third and fourth moments by

$$\begin{aligned} U_{jkl} &= \sqrt{n} \text{ave}_i \{z_{ij} z_{ik} z_{il}\}, \\ V_{jk} &= \sqrt{n} \text{ave}_i \{z_{ij}^2 z_{ik}^2 - 1\} \end{aligned}$$

and

$$V_{jj} = \sqrt{n} \text{ave}_i \{z_{ij}^4 - 3\}.$$

The limiting distributions of the standardized third and fourth moments are given by the following theorem. The results easily follow from Lemmas 7.1 and 7.2 stated and proven in the Appendix.

**Theorem 4.1.** *Under multinormality, the limiting distribution of the vector of all possible third moments  $U_{jkl}$  ( $j < k < l$ ),  $U_{jjk}$  ( $j \neq k$ ) and  $U_{jjj}$  is a multivariate normal distribution with zero mean vector and limiting variances and covariances*

$$\begin{aligned} \text{Var}(U_{jkl}) &= 1 \\ \text{Var}(U_{jjk}) &= 2 + \kappa \\ \text{Var}(U_{jjj}) &= 6 + 9\kappa \\ \text{Cov}(U_{jjk}, U_{llk}) &= \kappa \\ \text{Cov}(U_{jjk}, U_{kkk}) &= 3\kappa \end{aligned}$$

where

$$\kappa = 1 + \frac{E[\gamma^2(r)]}{p} - 2 \frac{E[r^3 \gamma(r)]}{p(p+2)}.$$

All the other covariances are zero.

**Theorem 4.2.** *Under multinormality, the limiting distribution of the vector of fourth moments  $V_{jk}$  ( $j \neq k$ ) and  $V_{jj}$  is a multivariate normal distribution with zero mean vector and limiting variances and covariances*

$$\begin{aligned} \text{Var}(V_{jk}) &= 4 + \tau_1 + 2\tau_2 \\ \text{Var}(V_{jj}) &= 24 + 9\tau_1 + 36\tau_2 \\ \text{Cov}(V_{jk}, V_{jl}) &= \tau_1 + \tau_2 \\ \text{Cov}(V_{jk}, V_{ml}) &= \tau_1 \\ \text{Cov}(V_{jj}, V_{kk}) &= 9\tau_1 \\ \text{Cov}(V_{jk}, V_{jj}) &= 3\tau_1 + 6\tau_2 \\ \text{Cov}(V_{jk}, V_{ll}) &= 3\tau_1 \end{aligned}$$

where

$$\tau_1 = 4 \left[ \frac{E[\alpha^2(r)]}{p(p+2)} - 2 \frac{E[\alpha(r)\beta(r)]}{p} + E[\beta^2(r)] - \frac{E[\alpha(r)r^4]}{p(p+2)(p+4)} + \frac{E[\beta(r)r^4]}{p(p+2)} \right]$$

and

$$\tau_2 = 2 + 2 \frac{E[\alpha^2(r)]}{p(p+2)} - 4 \frac{E[\alpha(r)r^4]}{p(p+2)(p+4)}.$$

We close this section with some discussion on the implications of the above results. First note that, if the regular mean vector and sample covariance matrix are used in the standardization, then simply

$$\gamma(r) = r, \quad \alpha(r) = r^2 \quad \text{and} \quad \beta(r) = 1$$

which gives

$$\kappa = \tau_1 = \tau_2 = 0$$

and all the third moments and fourth moments are asymptotically mutually independent, respectively. For other location and scatter statistics this is not necessarily true and also the limiting variances may vary. This means that the limiting distributions and efficiencies of  $b_{1,p}$  and  $b_{2,p}$  may drastically change. The limiting distribution of  $nb_{1,p}/6$  is a weighted sum of chi-square variables with one degree of freedom which makes the calculation of the approximate  $p$ -value less practical. The limiting distribution of  $\sqrt{n}(b_{2,p} - p(p+2))$  is still normal with mean value zero, but its limiting variance depends on the chosen scatter matrix. In the last two section we illustrate the impact of the choice of the location vector and scatter matrix estimate on the limiting distribution and efficiency in the bivariate case. The extension to the general  $p$ -variate case is straightforward.

## 5. Tests for bivariate normality

Assume now that  $\mathbf{x}_1, \dots, \mathbf{x}_n$  is a random sample from an unknown bivariate distribution and we wish to test the null hypothesis  $H_0$  of bivariate normality. For

limiting power considerations we use the contaminated bivariate normal distribution, the so called Tukey distribution: We say that the distribution of  $\mathbf{x}$  is a bivariate Tukey distribution  $T(\Delta, \boldsymbol{\mu}, \sigma)$  if the pdf of  $\mathbf{x}$  is

$$f(\mathbf{x}) = (1 - \Delta)\phi(\mathbf{x}) + \Delta\sigma^{-2}\phi((\mathbf{x} - \boldsymbol{\mu})/\sigma),$$

where  $\phi(\mathbf{x})$  is the pdf of  $N_2(\mathbf{0}, I_2)$  distribution and  $\Delta \in [0, 1]$ . Alternative sequences  $H_1^n : T(\Delta, \boldsymbol{\mu}, 1)$  and  $H_2^n : T(\Delta, \mathbf{0}, \sigma)$  with  $\Delta = \delta/\sqrt{n}$ ,  $r = \|\boldsymbol{\mu}\| > 0$  and  $\sigma > 1$  are then used for the comparisons of the skewness and kurtosis test statistics, respectively.

Write now

$$\mathbf{T}_1 = (U_{112}, U_{221}, U_{111}, U_{222})^T$$

and

$$\mathbf{T}_2 = (V_{11}, V_{22}, V_{12})^T$$

for the vectors of the standardized third and fourth moments. Using the results in Section 4, one directly gets the following results.

**Corollary 5.1.** *Under  $H_0$ , the limiting distribution of  $\sqrt{n}\mathbf{T}_1$  is a 4-variate normal distribution with zero mean vector and covariance matrix*

$$\Omega_1 = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 6 & 0 \\ 0 & 0 & 0 & 6 \end{pmatrix} + \kappa \begin{pmatrix} 1 & 0 & 0 & 3 \\ 0 & 1 & 3 & 0 \\ 0 & 3 & 9 & 0 \\ 3 & 0 & 0 & 9 \end{pmatrix}.$$

**Corollary 5.2.** *Under  $H_0$ , the limiting distribution of  $\sqrt{n}\mathbf{T}_2$  is a 3-variate normal distribution with zero mean vector and covariance matrix*

$$\Omega_2 = \begin{pmatrix} 24 & 0 & 0 \\ 0 & 24 & 0 \\ 0 & 0 & 4 \end{pmatrix} + \tau_1 \begin{pmatrix} 9 & 9 & 3 \\ 9 & 9 & 3 \\ 3 & 3 & 1 \end{pmatrix} + \tau_2 \begin{pmatrix} 36 & 0 & 6 \\ 0 & 36 & 6 \\ 6 & 6 & 2 \end{pmatrix}.$$

It is now possible to construct the limiting distributions of the affine invariant test statistics  $b_{1,2}$  and  $b_{2,2}$ . The comparison of the effect of different standardizations is possible but not easy, however, as the limiting distributions may be of different type. This is avoided if one uses the following test statistic versions:

**Definition 5.3.** Modified Mardia's skewness and kurtosis statistics in the bivariate case are

$$b_{1,2}^* = \mathbf{T}_1^T \Omega_1^{-1} \mathbf{T}_1 \quad \text{and} \quad b_{2,2}^* = (1, 1, 2) \mathbf{T}_2.$$

Note that, if the sample mean and covariance matrix are used, the definition gives  $b_{1,2}^* = b_{1,2}/6$  and  $b_{2,2}^* = b_{2,2} - 8$ , respectively. In the former case, with general  $\hat{\boldsymbol{\mu}}$  and  $\hat{\Sigma}$ , the affine equivariance property is lost.

The alternative to the regular sample mean vector and sample covariance matrix used in this example is the affine equivariant Oja median (Oja, 1983) and the related scatter matrix estimate based on the Oja sign covariance matrix (SCM).

See Visuri et al. (2000) and Ollila et al. (2003b) for the SCM. Note also that the scatter matrix estimate based on the Oja SCM is asymptotically equivalent with the zonoid covariance matrix (ZCM) and can be seen as an affine equivariant extension on mean deviation. For the definition and use of the ZCM, see Koshevoy et al (2003). We chose these statistics as their influence functions are relatively simple at the  $N_2(\mathbf{0}, I_2)$  case, the influence functions are given by

$$\gamma(r) = 2\sqrt{\frac{2}{\pi}}, \quad \alpha(r) = 2 \left[ 2\sqrt{\frac{2}{\pi}} r - 1 \right] \quad \text{and} \quad \beta(r) = 1.$$

The constants needed for the asymptotic variances and covariances are then

$$\kappa = \frac{4}{\pi} - \frac{1}{2}, \quad \tau_1 = \frac{32}{\pi} - \frac{29}{3} \quad \text{and} \quad \tau_2 = \frac{16}{\pi} - \frac{14}{3}.$$

Consider first the alternative contiguous sequence  $H_1^n$  with  $\boldsymbol{\mu} = (r, 0, \dots, 0)^T$ . Using classical LeCam's third lemma one then easily sees that, under the alternative sequence  $H_1^n$ , the limiting distribution of  $nb_{1,2}^*$  is a noncentral chi-square distribution with 4 degrees of freedom and noncentrality parameter  $\boldsymbol{\delta}_1^T \Omega_1^{-1} \boldsymbol{\delta}_1$ , where  $\boldsymbol{\delta}_1$  depends on the location estimate used in the standardization. In the regular case (sample mean vector),

$$\boldsymbol{\delta}_1 = (0, 0, r^3, 0)^T$$

and the Oja median gives

$$\boldsymbol{\delta}_1 = (0, r - q(r)2\sqrt{2/\pi}, r^3 + 3r - q(r)6\sqrt{2/\pi}, 0)^T.$$

The function  $q(r) = \|R(\boldsymbol{\mu})\|$  is easily computable and defined through the so called spatial rank function

$$R(\boldsymbol{\mu}) = E(\|\boldsymbol{\mu} - \mathbf{x}\|^{-1}(\boldsymbol{\mu} - \mathbf{x}))$$

with  $\mathbf{x} \sim N_2(\mathbf{0}, I_2)$ . See Möttönen and Oja (1995).

Next we consider the sequence  $H_2^n$ . Then

$$Var(b_{2,2}^*) = (1, 1, 2) \Omega_2 \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$$

and under  $H_2^n$  the limiting distribution of  $n(b_{2,2}^*)^2 / Var(b_{2,2}^*)$  is a noncentral chi-square distribution with one degree of freedom and noncentrality parameter  $[(1, 1, 2)\boldsymbol{\delta}_2]^2 / Var(b_{2,2}^*)$ , where  $\boldsymbol{\delta}_2$  in this case depends on the scatter estimate used in the standardization. In the regular case (sample covariance matrix) we get

$$\boldsymbol{\delta}_2 = (3(\sigma^2 - 1)^2, 3(\sigma^2 - 1)^2, (\sigma^2 - 1)^2)^T$$

and the ZCM gives

$$\boldsymbol{\delta}_2 = (3(\sigma^4 - 1) - 12(\sigma - 1), 3(\sigma^4 - 1) - 12(\sigma - 1), (\sigma^4 - 1) - 4(\sigma - 1))^T$$

In Figure 1 the asymptotical relative efficiencies of the new tests (using the Oja median and the ZCM estimate) relative to the regular tests (using the sample mean and sample covariance matrix) are illustrated for different values of  $r$  and

for different values of  $\sigma$ . The AREs are then simply the ratios of the noncentrality parameters. The new tests seem to perform very well: The test statistics using robust standardization seem to perform better than the tests with regular location and scatter estimates.

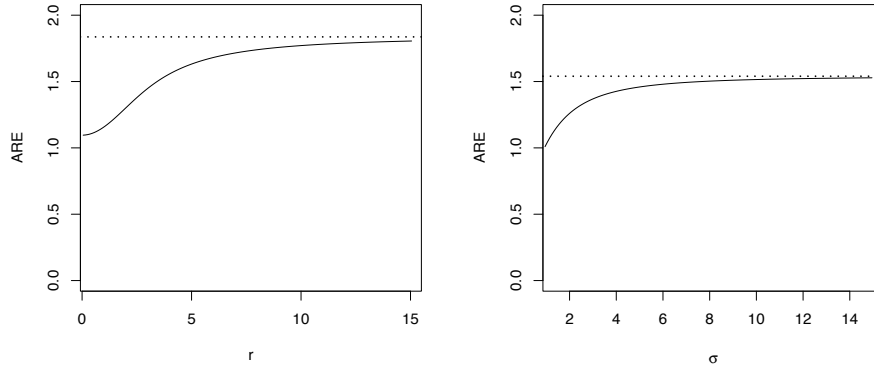


FIGURE 1. Limiting efficiencies of the tests based on the Oja median and the ZCM relative to the regular Mardia's tests for different values of  $r$  and  $\sigma$ . The left (right) panel illustrates the efficiency of  $b_{1,2}^*$  ( $b_{2,2}^*$ ).

## 6. A real data example

In our real data example we consider two bivariate data sets of  $n = 50$  observations and the combined data set ( $n = 100$ ). The data sets are shown and explained in Figure 2. We wish to test the null hypothesis of bivariate normality. Again the the skewness and kurtosis statistics to be compared are those based on the sample mean vector and sample covariance matrix (regular Mardia's  $b_{1,2}$  and  $b_{2,2}$ ) and those based on the Oja median and the ZCM (denoted by  $b_{1,2}^*$  and  $b_{2,2}^*$ ). The values of the test statistics and the approximate  $p$ -values were calculated for the three data sets ('Boys', 'Girls', 'Combined'). See Table 1 for the results. As expected, the observed  $p$ -values were much smaller in the combined data case.

As discussed in the introduction, the statistics  $b_{1,2}$  and  $b_{2,2}$  compare the third and fourth central moments to the second ones (given by the sample covariance matrix). The statistics  $b_{1,2}^*$  and  $b_{2,2}^*$  use robust location vector and scatter matrix estimates in this comparison. One can then expect that, for data sets with outliers or for data sets coming from heavy tailed distributions, the comparison to robust estimates tends to produce higher values of the test statistics and smaller  $p$ -values.



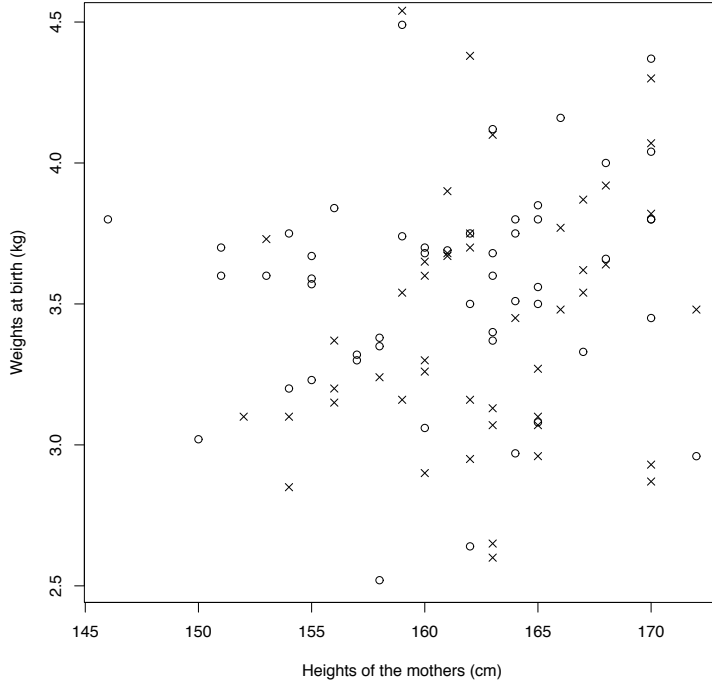


FIGURE 2. The height of the mother and the birth weight of the child measured on 50 boys who have siblings (o) and 50 girls (x) who do not have siblings.

TABLE 1. The observed values of the test statistics with corresponding  $p$ -values for the three data sets.

	Boys		Girls		Combined	
	Test statistic	p-value	Test statistic	p-value	Test statistic	p-value
$b_{1,2}$	3.034	0.552	3.376	0.497	9.242	0.055
$b_{1,2}^*$	5.152	0.272	2.336	0.674	11.257	0.024
$b_{2,2}$	0.016	0.900	0.282	0.596	2.275	0.132
$b_{2,2}^*$	0.834	0.361	1.152	0.283	5.166	0.023

In our examples for the analyses of bivariate data we used the Oja median and the ZCM to standardize the data vectors. In the general  $p$ -variate case (with large  $p$  and large  $n$ ) these estimates are, however, computationally highly intensive, and should be replaced by more practical robust location and scatter estimates.

## 7. Appendix: Auxiliary lemmas

The limiting distributions of the standardized third and fourth moments are given by the following linear approximations.

**Lemma 7.1.** *Let  $\mathbf{x}_1, \dots, \mathbf{x}_n$  be a random sample from  $N(\mathbf{0}, I_p)$ -distribution and let  $\mathbf{z}_i = \hat{\Sigma}^{-1/2}(\mathbf{x}_i - \hat{\boldsymbol{\mu}})$ ,  $i = 1, \dots, n$ , be the standardized observations. Write also  $r_i = \|\mathbf{x}_i\|$  and  $\mathbf{u}_i = \|\mathbf{x}_i\|^{-1}\mathbf{x}_i$ . Then*

$$\begin{aligned} \sqrt{n} \operatorname{ave}_i\{z_{ij}z_{ik}z_{il}\} &= \sqrt{n} \operatorname{ave}_i\{x_{ij}x_{ik}x_{il}\} + o_P(1) \\ &= \sqrt{n} \operatorname{ave}_i\{r_i^3 u_{ij}u_{ik}u_{il}\} + o_P(1) \quad \text{for } j < k < l, \\ \sqrt{n} \operatorname{ave}_i\{z_{ij}^2 z_{ik}\} &= \sqrt{n} \operatorname{ave}_i\{x_{ij}^2 x_{ik}\} - t_k + o_P(1) \\ &= \sqrt{n} \operatorname{ave}_i\{r_i^3 u_{ij}^2 u_{ik} - \gamma(r_i)u_{ik}\} + o_P(1) \quad \text{for } j \neq k \end{aligned}$$

and

$$\begin{aligned} \sqrt{n} \operatorname{ave}_i\{z_{ij}^3\} &= \sqrt{n} \operatorname{ave}_i\{x_{ij}^3\} - 3t_j + o_P(1) \\ &= \sqrt{n} \operatorname{ave}_i\{r_i^3 u_{ij}^3 - 3\gamma(r_i)u_{ij}\} + o_P(1). \end{aligned}$$

**Lemma 7.2.** *Let  $\mathbf{x}_1, \dots, \mathbf{x}_n$  be a random sample from  $N(\mathbf{0}, I_p)$ -distribution and let  $\mathbf{z}_i = \hat{\Sigma}^{-1/2}(\mathbf{x}_i - \hat{\boldsymbol{\mu}})$ ,  $i = 1, \dots, n$ , be the standardized observations. Again,  $r_i = \|\mathbf{x}_i\|$  and  $\mathbf{u}_i = \|\mathbf{x}_i\|^{-1}\mathbf{x}_i$ . Then for  $j \neq k$ ,*

$$\begin{aligned} \sqrt{n} \operatorname{ave}_i\{z_{ij}^2 z_{ik}^2 - 1\} &= \sqrt{n} \operatorname{ave}_i\{x_{ij}^2 x_{ik}^2 - 1\} - C_{jj} - C_{kk} + o_P(1) \\ &= \sqrt{n} \operatorname{ave}_i\{r_i^4 u_{ij}^2 u_{ik}^2 - 1 - \alpha(r_i)[u_{ij}^2 + u_{ik}^2] + 2\beta(r_i)\} + o_P(1) \end{aligned}$$

and

$$\begin{aligned} \sqrt{n} \operatorname{ave}_i\{z_{ij}^4 - 3\} &= \sqrt{n} \operatorname{ave}_i\{x_{ij}^4 - 3\} - 6C_{jj} + o_P(1) \\ &= \sqrt{n} \operatorname{ave}_i\{r_i^4 u_{ij}^4 - 3 - 6\alpha(r_i)u_{ij}^2 + 6\beta(r_i)\} + o_P(1). \end{aligned}$$

**Proofs of the lemmas:** Note first that

$$\mathbf{z}_i = \hat{\Sigma}^{-1/2}(\mathbf{x}_i - \hat{\boldsymbol{\mu}}) = \mathbf{x}_i - \frac{1}{\sqrt{n}}\mathbf{t} - \frac{1}{2\sqrt{n}}C\mathbf{x}_i + o(1/\sqrt{n}).$$

Thus

$$\begin{aligned} z_{ij} &= x_{ij} - \frac{1}{\sqrt{n}}t_j - \frac{1}{2\sqrt{n}}\sum_{r=1}^p C_{jr}x_{ir} + o(1/\sqrt{n}), \\ z_{ij}^2 &= x_{ij}^2 - \frac{2}{\sqrt{n}}t_j x_{ij} - \frac{1}{\sqrt{n}}\sum_{r=1}^p C_{jr}x_{ir}x_{ij} + o(1/\sqrt{n}), \\ z_{ij}^3 &= x_{ij}^3 - \frac{3}{\sqrt{n}}t_j x_{ij}^2 - \frac{3}{2\sqrt{n}}\sum_{r=1}^p C_{jr}x_{ir}x_{ij}^2 + o(1/\sqrt{n}) \end{aligned}$$

and

$$z_{ij}^4 = x_{ij}^4 - \frac{4}{\sqrt{n}} t_j x_{ij}^3 - \frac{2}{\sqrt{n}} \sum_{r=1}^p C_{jr} x_{ir} x_{ij}^3 + o(1/\sqrt{n}).$$

As the observations come from the  $N_p(\mathbf{0}, I_p)$  distribution (all the moments exist), the standard law of large numbers and central limit theorem together with Slutsky's lemma yield the result.

**Acknowledgements** We thank the referees for the critical and constructive comments. Finally, we wish to mention that, after writing the paper, we learned about the paper Croux et al. (2003) with a very similar but more general approach. Croux and his cowriters study the behaviour of the information matrix (IM) test when the maximum likelihood estimates are replaced by robust estimates. One of their examples is the modified Jarque-Bera test for univariate skewness and kurtosis. Their conclusion also is that the use of the robust estimates increases the power of the IM test.

## References

- [1] A. Bera and S. John, *Tests for multivariate normality with Pearson alternatives*. Commun. Statist. Theor. Meth. **12**(1) (1983), 103–117.
- [2] C. Croux, G. Dhaene and D. Hoorelbeke, *Testing the information matrix equality with robust estimators*. Submitted. (Website address: [www.econ.kuleuven.ac.be/public/NDBAE06/public.htm](http://www.econ.kuleuven.ac.be/public/NDBAE06/public.htm))
- [3] C. Croux and G. Haesbroeck, *Principal component analysis based on robust estimators of the covariance or correlation matrix: Influence functions and efficiencies*. Biometrika **87** (2000), 603–618.
- [4] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, P. J. and W.A. Stahel, *Robust Statistics: The Approach Based on Influence Functions*. Wiley, 1986.
- [5] G. Koshevoy, J. Mitten and H. Oja, *A scatter matrix estimate based on the zonotope*. Ann. Statist. **31** (2003), 1439–1459.
- [6] J. A. Koziol, *Probability plots for assessing multivariate normality*. Statistician **42** (1993), 161–173.
- [7] K. V. Mardia, *Measures of multivariate skewness and kurtosis with applications*. Biometrika **57** (1970), 519–530.
- [8] K. V. Mardia, *Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies*. Sankhyā, Ser B **36**(2) (1974), 115–128.
- [9] K. V. Mardia, *Tests of univariate and multivariate normality*. In: P.R. Krishnaiah (ed.), Handbook of Statistics, vol 1, 279–320, North Holland, 1980.
- [10] J. Möttönen and H. Oja, *Multivariate spatial sign and rank methods*. J. Nonparam. Statist. **5** (1995), 201–213.
- [11] H. Oja, *Descriptive statistics for multivariate distributions*. Statist. Probab. Lett. **1** (1983), 327–332.

- [12] E. Ollila, T.P. Hettmansperger and H. Oja, *Affine equivariant multivariate sign methods*. (2003a) Under revision.
- [13] E. Ollila, H. Oja and C. Croux, *The affine equivariant sign covariance matrix: Asymptotic behavior and efficiencies*. J. Multiv. Analysis (2003b), in print.
- [14] S. Visuri, V. Koivunen, and H. Oja, *Sign and rank covariance matrices*. J. Statist. Plann. Inference **91** (2000), 557–575.

University of Jyväskylä, Department of Mathematics and Statistics, P.O. Box 35  
(MaD), FIN-40014 University of Jyväskylä, Finland  
*E-mail address:* kankaine@maths.jyu.fi

University of Jyväskylä, Department of Mathematics and Statistics, P.O. Box 35  
(MaD), FIN-40014 University of Jyväskylä, Finland  
*E-mail address:* slahola@maths.jyu.fi

University of Jyväskylä, Department of Mathematics and Statistics, P.O. Box 35  
(MaD), FIN-40014 University of Jyväskylä, Finland  
*E-mail address:* ojahannu@maths.jyu.fi