MULTIVARIATE SKEWNESS AND KURTOSIS

by

Steven J. Schwager

Biometrics Unit, Cornell University, Ithaca, NY 14853

BU-802-M

Revised October, 1983

Abstract

Multivariate skewness and kurtosis are defined and their proper-
ties are discussed. Their use in testing for multivariate normality
and in robustness studies is described. Definitions given by Mardia,
Malkovich and Afifi, Isogai, and Oja are treated.

MULTIVARIATE SKEWNESS AND KURTOSIS

Multivariate skewness and kurtosis are generalizations of (univariate) skewness[*] and kurtosis[*], the standardized third and fourth moments, to multivariate distributions and samples. Other measures of univariate asymmetry, such as Pearson's (mean - mode)$/\sigma$, are also referred to as skewness, and can be similarly generalized.

Let F denote an arbitrary p-dimensional distribution, $\underset{\sim}{\mu}$ its $p \times 1$ mean vector, and $\underset{\sim}{\Sigma}$ its $p \times p$ covariance matrix. Let $\underset{\sim}{X}_1, \cdots, \underset{\sim}{X}_n$ denote a set of $p \times 1$ observations whose sample mean vector and covariance matrix are

$$\underset{\sim}{\bar{X}} = (1/n) \sum_{i=1}^{n} \underset{\sim}{X}_i \quad \text{and} \quad \underset{\sim}{S} = (1/n) \sum_{i=1}^{n} (\underset{\sim}{X}_i - \underset{\sim}{\bar{X}})(\underset{\sim}{X}_i - \underset{\sim}{\bar{X}})' .$$

In many of the computations below, n must exceed p .

Mardia [5,6] defined the multivariate skewness $\beta_{1,p}$ and kurtosis $\beta_{2,p}$ of distribution F as

$$\beta_{1,p} = E\{[(\underset{\sim}{X} - \underset{\sim}{\mu})'\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{Y} - \underset{\sim}{\mu})]^3\}$$

and

$$\beta_{2,p} = E\{[(\underset{\sim}{X} - \underset{\sim}{\mu})'\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{X} - \underset{\sim}{\mu})]^2\} ,$$

where $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ are independent $p \times 1$ random vectors with this distribution. He defined the multivariate sample skewness $b_{1,p}$ and kurtosis $b_{2,p}$ of the set of observations $\underset{\sim}{X}_1, \cdots, \underset{\sim}{X}_n$ as

$$b_{1,p} = (1/n^2) \sum_{i,j=1}^{n} [(\underset{\sim}{X}_i - \bar{\underset{\sim}{X}})' \underset{\sim}{S}^{-1} (\underset{\sim}{X}_j - \bar{\underset{\sim}{X}})]^3$$

and

$$b_{2,p} = (1/n) \sum_{i=1}^{n} [(\underset{\sim}{X}_i - \bar{\underset{\sim}{X}})' \underset{\sim}{S}^{-1} (\underset{\sim}{X}_i - \bar{\underset{\sim}{X}})]^2 .$$

For any nonsingular $p \times p$ matrix $\underset{\sim}{A}$ and any $p \times 1$ vector $\underset{\sim}{D}$, $b_{1,p}$ and $b_{2,p}$ are invariant* under the affine transformation $\underset{\sim\sim}{A}\underset{\sim}{X} + \underset{\sim}{D}$ of the sample; $\beta_{1,p}$ and $\beta_{2,p}$ are also invariant under this transformation. When the dimension p is 1, $\beta_{1,p}$ and $b_{1,p}$ reduce to the squares of the usual univariate population skewness $\sqrt{\beta_1}$ and sample skewness $\sqrt{b_1}$. In addition, $\beta_{2,p}$ and $b_{2,p}$ reduce to the usual univariate population kurtosis β_2 and sample kurtosis b_2. When $p = 2$ and $\underset{\sim}{\Sigma} = \underset{\sim}{I}$,

$$\beta_{1,p} = \mu_{30}^2 + \mu_{03}^2 + 3\mu_{12}^2 + 3\mu_{21}^2 ,$$

and

$$\beta_{2,p} = \mu_{04} + \mu_{40} + 2\mu_{22} .$$

Similar but more complex expressions are available for $p > 2$, $\underset{\sim}{\Sigma} \neq \underset{\sim}{I}$ [see 5].

The skewness of any distribution symmetric about its mean is $\beta_{1,p} = 0$. Thus the p-dimensional multivariate normal $N(\underset{\sim}{\mu}, \underset{\sim}{\Sigma})$ distribution has skewness $\beta_{1,p} = 0$. For a random sample from this distribution, the statistic $nb_{1,p}/6$ has an asymptotic chi-square distribution* with $p(p + 1)(p + 2)/6$ degrees of freedom. An improved version of this result

follows from the mean of $b_{1,p}$ under normal sampling [6].

The p-dimensional multivariate normal $N(\underset{\sim}{\mu}, \underset{\sim}{\Sigma})$ distribution has kurtosis $\beta_{2,p} = p(p + 2)$. For a random sample from this distribution, the statistic

$$[b_{2,p} - p(p + 2)(n - 1)/(n + 1)]/[8p(p + 2)/n]^{\frac{1}{2}}$$

is asymptotically normal $(0,1)$. (See ASYMPTOTIC NORMALITY.) An improved version of this result follows from the mean and variance of $b_{2,p}$ under normal sampling [6].

Mardia [5,6,7] advocated using his multivariate sample skewness and multivariate sample kurtosis to test for normality. A test based on skewness is given by rejecting the hypothesis of multivariate normality if $b_{1,p}$ is very large. A test based on kurtosis may be performed by rejecting the hypothesis of multivariate normality if $b_{2,p}$ is either very large or very small. To perform these tests, tables of critical points of the distributions of $b_{1,p}$ and $b_{2,p}$ under normal sampling are necessary for small to moderately large values of n. Tables for p = 2 and selected values of n from 10 to 5000, produced by Monte Carlo simulations* and smoothing, appear in Mardia [6], where recommendations for the case of p > 2 are also found. For extremely large n, critical points of $b_{1,p}$ and $b_{2,p}$ can be approximated from their asymptotic behavior. Tests of normality that use both

$b_{1,p}$ and $b_{2,p}$ have also been suggested. (See MARDIA'S TEST OF MULTINORMALITY.) Schwager and Margolin [10] showed that rejecting the null hypothesis whenever $b_{2,p}$ is sufficiently large gives the locally best* invariant test of H_0: the data are a multivariate normal random sample against H_1: there are some outliers resulting from mean slippage*.

The magnitude of the multivariate skewness $\beta_{1,p}$ and the extent to which the multivariate kurtosis $\beta_{2,p}$ differs from $p(p + 2)$ are measures of the nonnormality of a distribution. These can be useful in robustness studies. Nonnormality reflected by $\beta_{1,p}$ affects the size of Hotelling's T^2 test*, while nonnormality reflected by $\beta_{2,p}$ does not appear to have much impact on the size of this test. In contrast, nonnormality reflected by $\beta_{2,p}$ affects the size of the normal theory likelihood ratio test* for equal covariance matrices in several populations, which does not seem to be influenced much by nonnormality reflected in $\beta_{1,p}$ [5,6,7].

Algorithms for computing $b_{1,p}$ and $b_{2,p}$ were given by Mardia and Zemroch [8]. Gnanadesikan [1, Ch. 5] used $b_{1,p}$ and $b_{2,p}$ in analyzing several data sets.

Malkovich and Afifi [4] introduced different definitions of multivariate skewness and kurtosis, based on Roy's union-intersection principle*. If X has distribution F, then for any nonzero $p \times 1$ vector C, the scalar variable

$C'X$ has squared skewness

$$\beta_1(C) = \{E[(C'X - C'\mu)^3]\}^2/(C'\Sigma C)^3 .$$

The multivariate skewness of the distribution of X is defined as

$$\beta_1^M = \max_C \beta_1(C) ,$$

the largest squared skewness produced by any projection of the p-dimensional distribution onto a line. Similarly, the scalar variable $C'X$ has kurtosis

$$\beta_2(C) = E[(C'X - C'\mu)^4]/(C'\Sigma C)^2 .$$

The multivariate kurtosis of the distribution of X is defined as

$$\beta_2^M = \max_C|\beta_2(C) - 3| ;$$

this is the greatest deviation from 3, the kurtosis of the univariate normal distribution, produced by any projection of the p-dimensional distribution onto a line. The multivariate normal $N(\mu,\Sigma)$ distribution has $\beta_1^M = 0$ and $\beta_2^M = 0$, since every scalar variable $C'X$ is univariate normal, so $\beta_1(C) = 0$ and $\beta_2(C) = 3$ for every C.

For a sample $X_1, \cdots, X_n$ and any nonzero p X 1 vector C, the square of the sample skewness of the scalars $C'X_1, \cdots, C'X_n$ is

Steven J. Schwager

5

6

$$b_1(\underset{\sim}{C}) = n\left[\sum_{i=1}^{n}(\underset{\sim}{C}'\underset{\sim}{X}_i - \underset{\sim}{C}'\underset{\sim}{\bar{X}})^3\right]^2 \bigg/ \left[\sum_{i=1}^{n}(\underset{\sim}{C}'\underset{\sim}{X}_i - \underset{\sim}{C}'\underset{\sim}{\bar{X}})^2\right]^3 .$$

The multivariate sample skewness of $\underset{\sim}{X}_1, \cdots, \underset{\sim}{X}_n$ is [4]

$$b_1^M = \max_{\underset{\sim}{C}} b_1(\underset{\sim}{C}) .$$

Similarly, the sample kurtosis of $\underset{\sim}{C}'\underset{\sim}{X}_1, \cdots, \underset{\sim}{C}'\underset{\sim}{X}_n$ is

Steven J. Schwager

5

7

$$b_2(\underset{\sim}{C}) = n \sum_{i=1}^{n} (\underset{\sim}{C}'\underset{\sim}{X_i} - \underset{\sim}{C}'\underset{\sim}{\bar{X}})^4 \Bigg/ \left[\sum_{i=1}^{n} (\underset{\sim}{C}'\underset{\sim}{X_i} - \underset{\sim}{C}'\underset{\sim}{\bar{X}})^2 \right]^2 .$$

A union-intersection test of multivariate normality based on kurtosis is given by rejecting the hypothesis of normal random sampling whenever $b_2(\underset{\sim}{C})$ is far from 3 for any $\underset{\sim}{C}$. The multivariate sample kurtosis of $\underset{\sim}{X_1}, \cdots, \underset{\sim}{X_n}$ is [4]

$$b_2^M = \max_{\underset{\sim}{C}} \left| b_2(\underset{\sim}{C}) - K \right| ,$$

where the constant K is chosen to equalize, under the hypothesis of multivariate normality, the probabilities of rejecting this hypothesis because $\min_{\underset{\sim}{C}} b_2(\underset{\sim}{C})$ is very small and because $\max_{\underset{\sim}{C}} b_2(\underset{\sim}{C})$ is very large. As n increases, $K \to 3$.

For any p, b_1^M and b_2^M are invariant under nonsingular affine transformations $\underset{\sim\sim}{AX} + \underset{\sim}{D}$, as are β_1^M and β_2^M. When the dimension p is 1, β_1^M and b_1^M reduce to the squares β_1 and b_1 of the usual population and sample skewness. Also, $\beta_2^M = |\beta_2 - 3|$ and $b_2^M = |b_2 - K|$, where β_2 and b_2 are the usual population and sample kurtosis.

Malkovich and Afifi proposed using their multivariate sample skewness and kurtosis to test for multivariate normality. A union-intersection test based on skewness is given by rejecting the null hypothesis of normal random sampling whenever b_1^M is very large. The analogous test based on kurtosis was discussed in defining b_2^M. The

maximization and evaluation of K required to calculate b_1^M and b_2^M involve computations whose difficulty increases with p.

Isogai [2] extended Pearson's measure of univariate skewness, (mean - mode)$/\sigma$, to multivariate distributions and samples. He defined a measure τ_p of the multivariate skewness of the distribution F,

$$\tau_p = (\underset{\sim}{\mu} - \underset{\sim}{\theta})'\underset{\sim}{\omega}^{-1}(\underset{\sim}{\Sigma})(\underset{\sim}{\mu} - \underset{\sim}{\theta}) ,$$

where $\underset{\sim}{\theta}$ is the mode of F and $\underset{\sim}{\omega}(\underset{\sim}{\Sigma})$ is a specified p X p function of $\underset{\sim}{\Sigma}$, possibly equal to $\underset{\sim}{\Sigma}$ itself. The multivariate sample skewness of $\underset{\sim}{X}_1, \cdots, \underset{\sim}{X}_n$ is

$$t_p = (\bar{\underset{\sim}{X}} - \hat{\underset{\sim}{\theta}})'\underset{\sim}{\omega}^{-1}(\underset{\sim}{S})(\bar{\underset{\sim}{X}} - \hat{\underset{\sim}{\theta}}) ,$$

where the sample mode $\hat{\underset{\sim}{\theta}}$ is obtained by density estimation* with an appropriate kernel function. When $\underset{\sim}{X}_1, \cdots, \underset{\sim}{X}_n$ are a random sample from a multivariate normal distribution, t_p is asymptotically distributed as a linear combination of p independent X_1^2 variables. Isogai suggested using t_p to test for multivariate normality.

Neither t_p nor τ_p is invariant under nonsingular affine transformations. If the dimension p is 1 and ω is the identity function, then τ_p reduces to the square of Pearson's measure of population skewness, and t_p to its

sample analogue. These are not directly related to β_1 and b_1 in general; however, a relationship among τ_p $(p = 1)$, β_1, and β_2 holds for any distribution in the Pearson system[*] [see 3, pp. 85, 149].

Oja [9] defined multivariate skewness and kurtosis by considering the volume of the simplex in p-dimensional space determined by $p + 1$ points $\underset{\sim}{x}_1, \cdots, \underset{\sim}{x}_{p+1}$. Let $\underset{\sim}{x}_i = (x_{i1}, x_{i2}, \cdots, x_{ip})'$ for $i = 1, \cdots, p + 1$. The volume of this simplex is

$$\Delta(\underset{\sim}{x}_1, \cdots, \underset{\sim}{x}_{p+1}) = \frac{1}{p!} \left| \det \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{p+1,1} \\ \vdots & \vdots & & \vdots \\ x_{1p} & x_{2p} & \cdots & x_{p+1,p} \end{bmatrix} \right| .$$

A $p \times 1$ measure of location for the distribution F is given by $\underset{\sim}{\mu}_\alpha$ $(0 < \alpha < \infty)$ satisfying

$$E\{ \Delta(\underset{\sim}{X}_1, \cdots, \underset{\sim}{X}_p, \underset{\sim}{\mu}_\alpha)^\alpha \} = \min_{\underset{\sim}{\lambda}} E\{ \Delta(\underset{\sim}{X}_1, \cdots, \underset{\sim}{X}_p, \underset{\sim}{\lambda})^\alpha \} ,$$

where $\underset{\sim}{X}_1, \cdots, \underset{\sim}{X}_p$ are independent $p \times 1$ random vectors with this distribution. The mean $\underset{\sim}{\mu}$ of F equals $\underset{\sim}{\mu}_2$, and $\underset{\sim}{\mu}_1$ can be used to define the generalized median of F. Oja defined the distribution's multivariate skewness as

$$\eta_p = E\{\Delta(\underset{\sim}{X}_1, \cdots, \underset{\sim}{X}_{p-1}, \underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2)\} / (E\{\Delta(\underset{\sim}{X}_1, \cdots, \underset{\sim}{X}_p, \underset{\sim}{\mu}_2)^2\})^{\frac{1}{3}}$$

and its multivariate kurtosis as

$$\beta^{*}_{2,p} = E\{\Delta(\underset{\sim}{X}_1, \cdots, \underset{\sim}{X}_p, \underset{\sim}{\mu}_2)^4\} / (E\{\Delta(\underset{\sim}{X}_1, \cdots, \underset{\sim}{X}_p, \underset{\sim}{\mu}_2)^2\})^2 \ ,$$

where $\underset{\sim}{X}_1, \cdots, \underset{\sim}{X}_p$ are independent vectors with this distribution.

For a sample $\underset{\sim}{X}_1, \cdots, \underset{\sim}{X}_n$, any $\underset{\sim}{\mu}_\alpha$ $(0 < \alpha < \infty)$ may be estimated by solving the equation

$$\sum \left[\Delta\left(\underset{\sim}{X}_{i_1}, \cdots, \underset{\sim}{X}_{i_p}, \widehat{\underset{\sim}{\mu}}_\alpha\right) \right]^\alpha = \min_{\underset{\sim}{\lambda}} \sum \left[\Delta\left(\underset{\sim}{X}_{i_1}, \cdots, \underset{\sim}{X}_{i_p}, \underset{\sim}{\lambda}\right) \right]^\alpha \ ,$$

where summation is over $1 \le i_1 < i_2 < \cdots < i_p \le n$. This gives the sample mean $\bar{\underset{\sim}{X}}$ as $\widehat{\underset{\sim}{\mu}}_2$. The sample median $\widehat{\underset{\sim}{\mu}}_1$ may be a single point, but may be a convex set from which the median can be chosen. Oja defined the multivariate sample skewness of $\underset{\sim}{X}_1, \cdots, \underset{\sim}{X}_n$ as

$$h_p = \binom{n}{p-1}^{-1} \sum \left[\Delta\left(\underset{\sim}{X}_{i_1}, \cdots, \underset{\sim}{X}_{i_{p-1}}, \widehat{\underset{\sim}{\mu}}_1, \bar{\underset{\sim}{X}}\right) \right] \Big/ v_2^{\frac{1}{2}}$$

and the multivariate sample kurtosis as

$$b^{*}_{2,p} = v_4 / v_2^2 \ ,$$

where

$$v_j = \binom{n}{p}^{-1} \sum \left[\Delta\left(\underset{\sim}{X}_{i_1}, \cdots, \underset{\sim}{X}_{i_p}, \bar{\underset{\sim}{X}}\right) \right]^j \ ,$$

with summation over $1 \le i_1 < i_2 < \cdots < i_{p-1} \le n$ in the

numerator of h_p and over $1 \le i_1 < \cdots < i_p \le n$ for each v_j.

For any dimension p, h_p and $b^*_{2,p}$ are invariant under

nonsingular affine transformations, as are η_p and $\beta^*_{2,p}$.

When p is 1, $\beta^*_{2,p} = \beta_2$ and $b^*_{2,p} = b_2$; in addition, η_p re-

duces to $|\mu - \mu_1|/\sigma$, where μ, μ_1, and σ are the mean,

median, and standard deviation of the distribution F, and

h_p reduces to the sample analogue of this quantity. These

are not directly related to β_1 and b_1.

References

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

[2] Isogai, T. (1982). Ann. Inst. Statist. Math., 34, A, 531-541.

[3] Kendall, M. G. and Stuart, A. (1969). The Advanced Theory of Statistics, Vol. 1, 3rd ed. Hafner, New York.

[4] Malkovich, J. F. and Afifi, A. A. (1973). J. Amer. Statist. Assoc., 68, 176-179.

[5] Mardia, K. V. (1970). Biometrika, 57, 519-530.

[6] Mardia, K. V. (1974). Sankhyā B, 36, 115-128.

[7] Mardia, K. V. (1975). Appl. Statist., 24, 163-171.

[8] Mardia, K. V. and Zemroch, P. J. (1975). Appl. Statist., 24, 262-265.

Steven J. Schwager

Encyclo. Statist. Sci.

5

12

[9] Oja, H. (   ). Statist. and Prob. Letters,

[10] Schwager, S. J. and Margolin, B. H. (1982). Ann.

Statist., 10, 943-954.

(DEPARTURES FROM NORMALITY, TESTS FOR

 KURTOSIS

 SKEWNESS)