*Sankhyā : The Indian Journal of Statistics*
1974, Volume **36**, Series B, Pt. 2, pp. 115–128.

# APPLICATIONS OF SOME MEASURES OF MULTIVARIATE SKEWNESS AND KURTOSIS IN TESTING NORMALITY AND ROBUSTNESS STUDIES

*By* K. V. MARDIA*
*University of Hull*

*SUMMARY.* Some measures of multivariate skewness and kurtosis have been proposed by Mardia (1970). For further theoretical investigations and applications, alternative forms of these measures are obtained. One of these forms is convenient for computer programming and, incidentally, provides a simpler proof of the invariance property of these measures. For mixtures of multivariate normal distributions, the measures are studied in relation to a measure of non-normality of Day (1969). Our investigation together with those of Hopkins and Clay (1963) indicates that the size of the normal theory tests of covariance matrices is extremely sensitive to kurtosis. A method to derive exact moments of these measures for samples from a multivariate normal population is developed; our approach is somewhat similar to Geary (1933) for the univariate case rather than Fisher (1930) who first solved this problem. Some suitable small sample approximations to the null distributions of the measures are then derived. Monte Carlo studies and these approximations are used to calculate critical values of a test of multivariate normality. A numerical example with small sample size is given.

## 1. Introduction

Let $\boldsymbol{X} = (X_1, \ldots, X_p)'$ be a random vector with mean vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)'$ and covariance matrix $\boldsymbol{\Sigma} = (\sigma_{rs})$. Let us write

$$\mu_{i_1 \cdots i_s}^{(r_1 \cdots r_s)} = E\left\{\prod_{k=1}^{s} (X_{r_k} - \mu_{r_k})^{i_k}\right\},$$

where $r_1, \ldots, r_s$ are integers and may not be different. Mardia (1970) has defined the measure of multivariate skewness as

$$\beta_{1,p} = \sum_{r,s,t=1}^{p} \sum_{r',s',t'=1}^{p} \sigma^{rr'}\sigma^{ss'}\sigma^{tt'}\mu_{111}^{(rst)}\,\mu_{111}^{(r's't')} \qquad \ldots \quad (1.1)$$

and the measure of multivariate kurtosis as

$$\beta_{2,p} = \sum_{r,s=1}^{p} \sum_{t,u=1}^{p} \sigma^{rs}\sigma^{tu}\mu_{1111}^{(rstu)}, \qquad \ldots \quad (1.2)$$

where $\boldsymbol{\Sigma}^{-1} = (\sigma^{rs})$. These measures appear naturally in certain studies on robustness of Hotelling's $T^2$.

Suppose that $\boldsymbol{X}_i' = (X_{1i}, \ldots, X_{pi})$, $i = 1, 2, \ldots, n$, are $n$ independent observations on $\boldsymbol{X}$. Let $\overline{\boldsymbol{X}}' = (\overline{X}_1, \ldots, \overline{X}_p)$ and $\boldsymbol{S} = (S_{rs})$ denote the sample mean

*Now at University of Leeds.

vector and covariance matrix respectively. For the sample, the measures of skewness and kurtosis corresponding to $\beta_{1,p}$ and $\beta_{2,p}$ are given by

$$b_{1,p} = \sum_{r,s,t}\sum_{r',s',t'} S^{rr'}S^{ss'}S^{tt'}M_{111}^{(rst)}M_{111}^{(r's't')} \qquad \dots \quad (1.3)$$

and

$$b_{2,p} = \sum_{r,s}\sum_{t,u} S^{rs}S^{tu}M_{1111}^{(rstu)} \qquad \dots \quad (1.4)$$

where

$$S^{-1} = (S^{rs}),\ M_{i_1 \dots i_s}^{(r_1 \dots r_s)} = \frac{1}{n}\sum_{i=1}^{n}\left\{\prod_{k=1}^{s}(X_{r_k i}-\bar{X}_{r_k})^{i_k}\right\}. \qquad \dots \quad (1.5)$$

Section 2 gives certain alternative forms of these measures and their simple uses. Section 3 deals with the effect of non-normality on the size of the normal theory tests of covariance matrices. Under the assumption of multivariate normality, Sections 4-6 obtain some exact moments of $b_{1,p}$ and $b_{2,p}$, and give new approximations to their distributions. Section 7 examines the old approximations (Mardia, 1970) and the new approximations by Monte Carlo trials, and gives critical values of a test of multivariate normality.

## 2. Alternative forms of the measures and some simple uses

2.1. *In terms of powers of certain bilinear forms.* We can express (1.1) as

$$\beta_{1,p} = E\{(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{\mu})\}^3, \qquad \dots \quad (2.1)$$

where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent and identical random vectors. Its sample counter-part is seen to be

$$b_{1,p} = \frac{1}{n^2}\sum_{i,j=1}^{n}\{(\boldsymbol{X}_i-\bar{\boldsymbol{X}})'\boldsymbol{S}^{-1}(\boldsymbol{X}_j-\bar{\boldsymbol{X}})\}^3. \qquad \dots \quad (2.2)$$

Further, (1.2) and (1.4) can be written as

$$\beta_{2,p} = E\{(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\}^2 \qquad \dots \quad (2.3)$$

and

$$b_{2,p} = \frac{1}{n}\sum_{i=1}^{n}\{(\boldsymbol{X}_i-\bar{\boldsymbol{X}})'\boldsymbol{S}^{-1}(\boldsymbol{X}_i-\bar{\boldsymbol{X}})\}^2. \qquad \dots \quad (2.4)$$

*Applications* (i) These expressions for $b_{1,p}$ and $b_{2,p}$ can easily be programmed by using the standard routines for inverting symmetric matrices and for evaluating bilinear forms. (Such an algorithm will appear in *J. Roy. Statist. Soc.*, **C**.)

(ii) Using the above expressions, we can prove their invariance property by much simpler argument than in Mardia (1970).

2.2. *Expressions in terms of cumulants and k-statistics.* We first obtain the population measures in terms of cumulants. Let $\kappa_{i_1 \cdots i_s}^{(r_1 \cdots r_s)}$ denote the cumulant of order $(i_1, \ldots, i_s)$ for the random variable $(X_{r_1}, \ldots, X_{r_s})$ where $r_1, r_2, \ldots, r_s$ are $s$ integers taking values $1, 2, \ldots, p$. The expressions corresponding to (1.1) and (1.2) for the cumulants are

$$\gamma_{1,p} = \Sigma\Sigma\ \kappa^{rr'}\kappa^{ss'}\kappa^{tt'}\kappa_{111}^{(rst)}\ \kappa_{111}^{(r's't')} \qquad \ldots \quad (2.5)$$

and

$$\gamma_{2,p} = \Sigma\Sigma\ \kappa^{rs}\kappa^{tu}\kappa_{1111}^{(rstu)}, \qquad \ldots \quad (2.6)$$

where

$$(\kappa_{11}^{(rs)})^{-1} = (\kappa^{rs}).$$

We now derive the relations between $\gamma_{1,p}$ and $\beta_{1,p}$ and between $\gamma_{2,p}$ and $\beta_{2,p}$. Following the usual method (Kendall and Stuart, 1969, pp. 83-84), it can be shown that

$$\kappa_{111}^{(rst)} = \mu_{111}^{(rst)}$$

so that from (1.1) and (2.5), we have

$$\gamma_{1,p} = \beta_{1,p}. \qquad \ldots \quad (2.7)$$

Similarly, it can be shown that

$$\kappa_{1111}^{(rstu)} = \mu_{1111}^{(rstu)} - (\sigma_{rs}\sigma_{tu} + \sigma_{rt}\sigma_{su} + \sigma_{ru}\sigma_{st}) \qquad \ldots \quad (2.8)$$

for all $r, s, t, u$ from 1 to $p$. On substituting (2.8) in (2.6) and using the results

$$\kappa^{rs} = \sigma^{rs},\ \Sigma\Sigma\sigma^{rs}\sigma^{tu}\sigma_{rs}\sigma_{tu} = (\Sigma\sigma^{rs}\sigma_{tu})^2 = p^2,\ \Sigma\Sigma\sigma^{rs}\sigma^{tu}\sigma_{rt}\sigma_{su} = p,$$

we have

$$\gamma_{2,p} = \beta_{2,p} - p(p+2). \qquad \ldots \quad (2.9)$$

Similarly, if $g_{1,p}$ and $g_{2,p}$ are the expressions (1.3) and (1.4) when the moments are replaced by the corresponding $k$-statistic, we have

$$g_{1,p} = n(n-1)b_{1,p}/(n-2)^2 \qquad \ldots \quad (2.10)$$

and

$$g_{2,p} = (n-1)\{(n+1)b_{2,p} - (n-1)p(p+2)\}/\{(n-2)(n-3)\}. \qquad \ldots \quad (2.11)$$

*Applications : Mixtures of multivariate normal distributions.* We first consider a mixture of two multivariate normal distributions with equal covariance matrices, i.e. the probability density function (p.d.f.) of $\boldsymbol{X}$ is given by

$$\lambda\phi(\boldsymbol{x}; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}) + \lambda'\phi(\boldsymbol{x}; \boldsymbol{\mu}_2, \boldsymbol{\Sigma}),\ -\infty < \boldsymbol{x} < \infty,\quad 0 < \lambda < 1, \qquad \ldots \quad (2.12)$$

where $\lambda' = 1-\lambda$ and $\phi(\cdot\,;\boldsymbol{\mu},\boldsymbol{\Sigma})$ is the p.d.f. corresponding to $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$. Day (1969, Section 8) has suggested that the Mahalanobis distance

$$\Delta = \{(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\}^{\frac{1}{2}}$$

can be taken as a measure of non-normality for this population and has given a test of normality based on an estimator of $\Delta$. We show for this population that $\gamma_{1,p}$ and $\gamma_{2,p}$ are some function of $\Delta$ so that our test of multivariate normality can also be used. By making suitable non-singular linear transformations (Kendall and Stuart, 1967, p. 229), we find that (2.12) reduces to

$$\{\lambda\phi(x_1-\Delta)+\lambda'\phi(x_1)\}\prod_{r=2}^{p}\phi(x_r),\quad -\infty < x_1,\ldots,x_p < \infty, \qquad \ldots \quad (2.13)$$

where $\phi(\cdot)$ is the p.d.f. of $N(0, 1)$. Since $\gamma_{1,p}$ and $\gamma_{2,p}$ are invariant under linear transformations, their values are unaltered. The joint cumulant generating function (c.g.f.) of (2.13) is

$$\frac{1}{2}\sum_{i=1}^{p} t_i^2+\log(\lambda'+\lambda e^{\Delta t_1}).$$

The relation of the second term with the c.g.f. of the point binomial distribution is obvious. Using this result, we find from (2.5) and (2.6),

$$\gamma_{1,p} = \{\lambda\lambda'(\lambda'-\lambda)\Delta^3\}^2/(1+\lambda\lambda'\Delta^2)^3 \qquad \ldots \quad (2.14)$$

and

$$\gamma_{2,p} = \lambda\lambda'(1-6\lambda\lambda')\Delta^4/(1+\lambda\lambda'\Delta^2)^2. \qquad \ldots \quad (2.15)$$

Hence $\gamma_{1,p}$ and $\gamma_{2,p}$ are some functions of $\Delta$. When $\Delta = 0$, $\gamma_{1,p}$ and $\gamma_{2,p}$ are zero as expected.

Let us now consider the case of a mixture of two normal distributions with unequal covariance matrices so that the p.d.f. of $X$ is given by

$$\lambda\Phi(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}_1)+\lambda'\Phi(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}_2),\quad -\infty < \boldsymbol{x} < \infty. \qquad \ldots \quad (2.16)$$

Under a suitable linear transformation (see Rao, 1965, pp. 37-38), it reduces to

$$\lambda\prod_{r=1}^{p}\{\phi(x_r/\sigma_r)/\sigma_r\}+\lambda'\left\{\prod_{r=1}^{p}\phi(x_r)\right\},\quad -\infty < \boldsymbol{x} < \infty. \qquad \ldots \quad (2.17)$$

From (2.5) and (2.6), it is found that

$$\gamma_{1,p} = 0$$

and

$$\gamma_{2,p} = \lambda\lambda'\left\{2\sum_{r=1}^{p} a_r^2+\left(\sum_{r=1}^{p} a_r\right)^2\right\},$$

where $a_r = (1-\sigma_r^2)(\lambda\lambda')^{\frac{1}{2}}/(\lambda\sigma_r^2+\lambda')$, $r = 1, 2, \ldots, p$. Obviously, $\gamma_{2,p}$ is non-negative. For $\sigma_1 = \sigma_2 = \ldots = \sigma_p = \sigma$, the distribution is completely symmetrical and

$$\gamma_{2,p} = p(p+2)(\lambda\lambda')^2(1-\sigma^2)^2/(\lambda\sigma^2+\lambda')^2. \quad \ldots \quad (2.18)$$

This result will be used in the next section.

## 3. Effect of non-normality on tests of covariance matrices

Mardia (1970) has shown that the size of the normal theory tests of mean vectors is more sensitive to $\gamma_{1,p}$ than $\gamma_{2,p}$. With the help of these measures, we now study the effect of non-normality on the size of the normal theory test of covariance matrices.

3.1. *The asymptotic case.* Let us first consider the $k$-sample problem. Suppose that $\boldsymbol{X}_m$, $m = 1, \ldots, k$ are $k$ $p\times 1$ random vectors with mean vectors $\boldsymbol{\mu}_m$ and covariance matrices $\boldsymbol{\Sigma}_m$ $(m = 1, 2, \ldots, k)$.

For testing

$$H_0 : \boldsymbol{\Sigma}_1 = \ldots = \boldsymbol{\Sigma}_m,$$

against the alternative that the $\boldsymbol{\Sigma}$'s are unspecified, the normal theory likelihood ratio criterion is

$$\lambda^{2/n} = \left\{ \prod_{m=1}^{k} |\boldsymbol{S}_m|^{r_m} \right\} \Big/ \left| \sum_{m=1}^{k} r_m \boldsymbol{S}_m \right|, \quad \ldots \quad (3.1)$$

where $r_m = n_m/n$, $n = \Sigma n_m$ and $\boldsymbol{S}_m$ is the sample covariance matrix for the $m$-th sample. It has been shown by Ito (1969) that under the null hypothesis, the random variable $n\{1-\lambda^{2/n}\}$ is asymptotically equivalent to the statistic $Y$ for any non-normal population where

$$Y = \frac{1}{2}\left[\sum_{m=1}^{k} tr\{(\boldsymbol{Z}_m\boldsymbol{\Sigma}_0^{-1})^2\} - tr\left\{\sum_{m=1}^{k} r_m^{\frac{1}{2}}\boldsymbol{Z}_m\boldsymbol{\Sigma}_0^{-1}\right\}^2\right], \quad \ldots \quad (3.2)$$

where $\boldsymbol{Z}_m = (z_{rs}^{(m)})$ and the $z_{rs}^{(m)}$'s are such that

$$E\{z_{rs}^{(m)}\, z_{tu}^{(m)}\} = \kappa_{rstu}^{(0)} + \sigma_{rt}^{(0)}\sigma_{su}^{(0)} + \sigma_{ru}^{(0)}\sigma_{st}^{(0)} \quad \ldots \quad (3.3)$$

and

$$E\{z_{rs}^{(m)} z_{tu}^{(m')}\} = 0, \quad m \neq m', \quad \ldots \quad (3.4)$$

with $\boldsymbol{\Sigma}_0 = (\sigma_{rs}^{(0)})$ a common covariance matrix and $\kappa_{rstu}^{(0)}$ is a cumulant of the fourth order of $\boldsymbol{X}_m$. With the help of (2.6) and (3.3), we have

$$E[tr\{(\boldsymbol{Z}_m\boldsymbol{\Sigma}_0^{-1})^2\}] = p(p+1)\left\{1+\frac{\gamma_{2,p}^{(0)}}{p(p+1)}\right\}$$

where $\gamma_{2,p}^{(0)}$ is the measure of kurtosis for the random vector $\boldsymbol{X}_m$. Further, on using (3.3) and (3.4), it is found that

$$E[tr\{(\Sigma r_m^{\frac{1}{2}} \boldsymbol{Z}_m \boldsymbol{\Sigma}_0^{-1})^2\}] = \Sigma r_m\{p(p+1)+\gamma_{2,p}^{(0)}\}.$$

Hence, from (3.2)

$$E(Y) = \frac{1}{2}(k-1)p(p+1)\left\{1+\frac{\gamma_{2,p}^{(0)}}{p(p+1)}\right\}. \qquad \dots \quad (3.5)$$

Consequently, it suggests that the size of the likelihood ratio test of equality of covariance matrices when the parent population is non-normal is seriously influenced by kurtosis and the effect becomes larger with $k$. Since (3.5) does not depend on $\gamma_{1,p}$, the effect of $\gamma_{1,p}$ on the size may not be serious.

Similarly, for the one-sample problem, it is found from Ito (1969) that under the null hypothesis

$$E(Y') = \frac{1}{2}p(p(p+1)\left\{1+\frac{\gamma_{2,p}^{(0)}}{p(p+1)}\right\}, \qquad \dots \quad (3.6)$$

where the statistic $Y'$ is asymptotically equivalent to $n(1-\lambda^{2/n})$, $\lambda$ is the likelihood ratio statistic for the one-sample normal theory problem. Hence, it is again expected that the size of the test is influenced by $\gamma_{2,p}$ but the effect of $\gamma_{1,p}$ may not be serious.

3.2. *The small sample-case.* Section 3.1 suggests that the size of the normal theory tests on covariance matrices would be seriously influenced by $\gamma_{2,p}$ but the effect of $\gamma_{1,p}$ may not be so serious. For $p = 1$, the result is well known (Box, 1953; Kendall and Stuart, 1967, pp. 466-68). We now show that the studies reported in the literature for $p > 1$ indicate that the above result holds for all values of $p$.

Consider the small sample investigation of Hopkins and Clay (1963). The authors have studied the effect on the $M$-statistic of Box for two samples from the mixture (2.17) with $p = 2$, $\lambda = .2$ and $\sigma_1 = \sigma_2 = \sigma$. For $n_1 = n_2$, the $M$-statistic can be seen to be a linear function of log $\lambda$ where $\lambda$ is given by (3.1) and so their investigations for $n_1 = n_2$ are applicable to the $\lambda$-test. They selected $\sigma = 2.5$ and $3.7$. Using (2.18), the corresponding values of $\gamma_{2,2}$ are found to be 1.3 and 2.6 respectively and $\gamma_{1,2} = 0$. Table 1 gives their results for the $\lambda$-test together with the corresponding results for Hotelling's $T^2$. The latter results are reproduced for comparisons. These results indicate that even for small sample, $\gamma_{2,2}$ has serious effect on the size of the likelihood ratio test for equality of covariance matrices although $\gamma_{1,2} = 0$.

TABLE 1. ACTUAL 100$\alpha$% LEVELS OF THE $\lambda$-TEST AND HOTELLING'S $T^2$ FOR TWO SAMPLES FROM THE MIXTURES OF BIVARIATE NORMAL DISTRIBUTIONS WITH 100$\alpha$% NOMINAL LEVELS
(based on tables of Hopkins and Clay, 1963)

| $n_1$ | $n_2$ | $\alpha$ | $\lambda$-test | | Hotelling's $T^2$ | |
|---|---|---|---|---|---|---|
| | | | $\gamma_{2,2} = 1.3$ | $\gamma_{2,2} = 2.6$ | $\gamma_{2,2} = 1.3$ | $\gamma_{2,2} = 2.6$ |
| 5 | 5 | 0.10 | 0.152 | 0.342 | 0.074 | 0.081 |
| | | 0.05 | 0.081 | 0.222 | 0.039 | 0.035 |
| 10 | 10 | 0.10 | 0.215 | 0.475 | 0.089 | 0.084 |
| | | 0.05 | 0.120 | 0.354 | 0.037 | 0.032 |
| 20 | 20 | 0.10 | 0.278 | 0.581 | 0.107 | 0.095 |
| | | 0.05 | 0.177 | 0.422 | 0.059 | 0.041 |

Next, consider the small sample investigation of Mardia (1971) for the two-sample criterion $V = tr\{\mathbf{S}_1(\mathbf{S}_1+\mathbf{S}_2)^{-1}\}$. In Section 7.4 of this reference, it is shown that the permutation distribution of $V$ can be approximated adequately by a beta distribution which depends only on a certain quantity $C_y$. It can be shown from equation (7.8) of Mardia (1971) that

$$E(C_y) = \frac{1}{p}\, \gamma_{2,p} + O(n^{-1}).$$

Hence, the above assertion also applies to the $V$-test.

## 4. A method for evaluations of moments of $b_{1,p}$ and $b_{2,p}$

Let $\boldsymbol{X}_i' = (X_{1i}, \ldots, X_{pi})$ $(i = 1, 2, \ldots, n)$ be a random sample from $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Since $b_{1,p}$ and $b_{2,p}$ are invariant under linear transformations, we assume without any loss of generality that $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \boldsymbol{I}$. From (2.2) and (2.4) we have

$$b_{1,p} = n \sum_{i,j=1}^{n} R_{ij}^3 \qquad \ldots \quad (4.1)$$

and

$$b_{2,p} = n \sum_{i=1}^{n} R_{ii}^2, \qquad \ldots \quad (4.2)$$

where

$$R_{ij} = (\boldsymbol{X}_i - \overline{\boldsymbol{X}})' \boldsymbol{M}^{-1} (\boldsymbol{X}_j - \overline{\boldsymbol{X}}) \qquad \ldots \quad (4.3)$$

and $$M = nS = \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})'. \quad \ldots \quad (4.4)$$

Consequently the moments of $b_{1,p}$ and $b_{2,p}$ can be determined with the help of the moments of $R_{ij}$. We now consider a method of obtaining the moments of $R_{ij}$.

4.1. *Some transformations.* Let us write $X_r^* = (X_{r1}, \ldots, X_{rn})'$ $(r = 1, 2, \ldots, p)$. We transform $X_r^*$ to $\xi_r^* = (\xi_{r1}, \ldots, \xi_{rn})'$ by the Helmert orthogonal transformation so that

$$\xi_{r,i-1} = \{(i-1)/i\}^{\frac{1}{2}}\left\{-X_{r,i} + \frac{1}{i-1}\sum_{k=1}^{i-1} X_{r,k}\right\}, \quad i = 2, 3, \ldots, n,$$

$$\xi_{r,n} = n^{\frac{1}{2}}\bar{X}_r,$$

where $\bar{X}_r$ is the mean of the observations on $X_r$. We have

$$M = \sum_{i=1}^{n-1} \xi_i \xi_i'$$

and $$X_i - \bar{X} = -a_i \xi_{i-1} + \sum_{k=i}^{n-1} b_k \xi_k, \quad i = 1, 2, \ldots, n,$$

with $$\xi_0 = \mathbf{0},\ \xi_n = -a_n \xi_{n-1},\ \xi_i' = (\xi_{1i}, \ldots, \xi_{pi}),$$

$$a_i = \{(i-1)/i\}^{\frac{1}{2}}, \quad b_i = \{i(i+1)\}^{-\frac{1}{2}}, \quad i = 1, 2, \ldots, n.$$

On using these results in (4.3) we find that

$$R_{ij} = \left(-a_i z_{i-1} + \sum_{k=i}^{n-1} b_k z_k\right)' \left(-a_j z_{j-1} + \sum_{k=j}^{n-1} b_k z_k\right), \quad \ldots \quad (4.5)$$

where $$z_i = T^{-1}\xi_i,\ i = 1, 2, \ldots, n, \text{ and } M = TT'. \quad \ldots \quad (4.6)$$

Hence the moments of $R_{ij}$ can be obtained from the moments of $z_1, \ldots, z_{n-1}$.

4.2. *Distribution of $Z_k$.* Let us write

$$S_1 = TT',\ S_2 = \sum_{i=k+1}^{n-1} \xi_i \xi_i', \quad V = (\xi_1, \ldots, \xi_k)$$

then we have $$S_1 = S_2 + VV'.$$

On using this identity, we find from Khatri (1959) that the probability density function (p.d.f.) of $Z_k$ is given by

$$g(Z_k; f, k, p) = C\,|I_p - Z_k Z_k'|^{\frac{1}{2}(f-p-k-1)},$$

where

$$C = \left[\prod_{r=1}^{p} \Gamma\left\{\frac{1}{2}(f-r+1)\right\}\right] \Big/ \left[\pi^{\frac{1}{2}kp} \prod_{r=1}^{p} \Gamma\left\{\frac{1}{2}(f-k-r+1)\right\}\right],$$

$f = n-1$, and $\boldsymbol{I}_p$ is the identity matrix of order $p \times p$. The distribution is a multivariate beta distribution. Khatri and Pillai (1967) have given an explicit transformation under which the density of $\boldsymbol{Z}_k$ reduces to the products of independent beta variables. Hence, the moments of $\boldsymbol{Z}_k$ can be obtained.

## 5. Approximations

Using the method of Section 4, it is found with the help of augmented symmetric function tables (David *et al.*, Tables 1, 6, 8, pp. 146-48) after some tedious algebra that

$$E(b_{1,p}) = \frac{p(p+2)}{(n+1)(n+3)} \{(n+1)(p+1)-6\}, \quad \ldots \quad (5.1)$$

$$E(b_{2,p}) = p(p+2)(n-1)/(n+1), \quad \ldots \quad (5.2)$$

and

$$\mathrm{var}(b_{2,p}) = \frac{8p(p+2)(n-3)}{(n+1)^2(n+3)(n+5)} (n-p-1)(n-p+1). \quad \ldots \quad (5.3)$$

For particular cases of $p = 1$ and $n \to \infty$, the above expressions reduce to the results given by Fisher (1930) and Mardia (1970), respectively. It has been shown in Mardia (1970) that

$$A = nb_{1,p}/6 \sim \chi^2_f, \; f = p(p+1)(p+2)/6. \quad \ldots \quad (5.4)$$

We can also consider the approximation

$$A' = nKb_{1,p}/6, \;\; K = (p+1)(n+1)(n+3)/[n\{(n+1)(p+1)-6\}] \quad \ldots \quad (5.5)$$

instead of $A$ so that $E(A') = f$ for all $n$.

On using the above expressions for $E(b_{2,p})$ and $\mathrm{var}(b_{2,p})$ and the central limit theorem, we observe that the statistic

$$B' = \frac{\{(n+1)b_{2,p}-p(p+2)(n-1)\}\{(n+3)(n+5)\}^{\frac{1}{2}}}{\{8p(p+2)(n-3)(n-p-1)(n-p+1)\}^{\frac{1}{2}}} \quad \ldots \quad (5.6)$$

is asymptotically distributed as $N(0, 1)$. Mardia (1970) has shown that

$$B = \{b_{2,p}-p(p+2)\}/\{8p(p+2)/n\}^{\frac{1}{2}} \quad \ldots \quad (5.7)$$

is asymptotically $N(0, 1)$. These approximations are examined in the next section.

3

## 6. Critical values of a test of multivariate normality

A test of multivariate normality has been proposed in Mardia (1970) which consists in testing $\beta_{1,p} = 0$ and $\beta_{2,p} = p(p+2)$ with the help of the test-statistics $b_{1,p}$ and $b_{2,p}$ respectively. For $p = 1$, critical values of these tests are already known (see Pearson and Hartley, 1966, pp. 207-208). Tables 2 and 3 give some selected critical values of the tests for $p = 2$ which were estimated as follows. First, the old approximations (5.4), (5.7) and the new approximations (5.5) and (5.6) to the null distributions of $b_{1,2}$ and $b_{2,2}$ were examined by Monte Carlo trials. It was found that these approximations were inadequate for moderately small values of $n$. Combinations of the two approximations provided crude approximations to 10% values which were used in working out at least 10% tails of the Monte Carlo distributions of $b_{1,2}$ and $b_{2,2}$ based on 10,000 samples for each value of $n$, for $n = 10(2)20, 25, 30(10)100$. For each value of $\alpha$, the Monte Carlo points were plotted against $N^{-1}$ for $b_{1,2}$ and $N^{-\frac{1}{2}}$ for $b_{2,2}$ together with the new and old approximations. For each value of $\alpha$, a smooth curve was drawn which approximated the points very closely. The resulting values from such curves are shown in Tables 2 and 3. For moderately large values of $n$, the new approximation should be used for $b_{1,2}$ while the old approximation should be used for $b_{2,2}$.

A close scrutiny of the curves together with the variance estimates of the points based on order statistics indicate that the significance points are accurate at least to the first two decimal places for $n \leqslant 20$, and accuracy increases with $n$.

For the Monte Carlo trials, random samples on variables $(X, Y)$ distributed as $N(\mathbf{0}, \boldsymbol{I})$ were drawn by using Box and Muller's method. A check on the suitability of the pseudo-random generator used in the investigation was made by comparing the observed and actual 2% and 10% points for $b_{1,1}$ and $b_{2,1}$ corresponding to the empirical marginal distributions of $(X, Y)$.

For $p > 2$, Monte Carlo studies together with the old and new approximations lead to the following recommendations in approximating the null distributions of $b_{1,p}$ and $b_{2,p}$ when $n \geqslant 50$. These approximations make the test overall conservative.

(1) For the upper 5% points of $b_{1,p}$, use the new approximation.

(2) For the lower 2.5% points of $b_{2,p}$, use $b_{2,p}$ distributed as normal with mean $p(p+2)(n+p+1)/n$ and variance $8p(p+2)/(n-1)$ when $50 \leqslant n \leqslant 400$, while use the old approximation when $n < 400$.

(3) For the upper 2.5% points of $b_{2,p}$, use the old approximation.

TABLE 2. CRITICAL VALUES OF $b_{1,2}$

| $n$ | $\alpha \to 0.001$ | 0.01 | 0.025 | 0.05 | 0.075 | 0.10 |
|---|---|---|---|---|---|---|
| 10 | 6.994 | 5.194 | 4.294 | 3.694 | 3.263 | 2.994 |
| 12 | 6.744 | 4.938 | 3.931 | 3.319 | 2.944 | 2.681 |
| 14 | 6.419 | 4.581 | 3.619 | 3.031 | 2.669 | 2.419 |
| 16 | 6.062 | 4.231 | 3.337 | 2.775 | 2.444 | 2.219 |
| 18 | 5.737 | 3.962 | 3.100 | 2.556 | 2.256 | 2.050 |
| 20 | 5.425 | 3.669 | 2.881 | 2.356 | 2.081 | 1.894 |
| 25 | 4.719 | 3.106 | 2.438 | 1.969 | 1.744 | 1.581 |
| 30 | 4.238 | 2.681 | 2.094 | 1.687 | 1.513 | 1.363 |
| 40 | 3.369 | 2.087 | 1.606 | 1.319 | 1.181 | 1.050 |
| 50 | 2.706 | 1.744 | 1.306 | 1.069 | 0.969 | 0.862 |
| 60 | 2.200 | 1.444 | 1.094 | 0.906 | 0.819 | 0.731 |
| 70 | 1.863 | 1.244 | 0.937 | 0.794 | 0.725 | 0.631 |
| 80 | 1.587 | 1.056 | 0.812 | 0.694 | 0.637 | 0.544 |
| 90 | 1.400 | 0.919 | 0.725 | 0.638 | 0.569 | 0.487 |
| 100 | 1.231 | 0.831 | 0.656 | 0.581 | 0.506 | 0.438 |
| 150 | 0.794 | 0.531 | 0.444 | 0.400 | 0.344 | 0.281 |
| 200 | 0.569 | 0.394 | 0.331 | 0.300 | 0.269 | 0.219 |
| 300 | 0.369 | 0.256 | 0.225 | 0.209 | 0.169 | 0.144 |
| 400 | 0.275 | 0.197 | 0.166 | 0.141 | 0.129 | 0.116 |
| 600 | 0.183 | 0.131 | 0.110 | 0.094 | 0.085 | 0.077 |
| 800 | 0.137 | 0.099 | 0.083 | 0.071 | 0.064 | 0.058 |
| 1000 | 0.110 | 0.079 | 0.066 | 0.057 | 0.051 | 0.046 |
| 1500 | 0.074 | 0.053 | 0.044 | 0.038 | 0.034 | 0.031 |
| 2500 | 0.044 | 0.032 | 0.027 | 0.023 | 0.021 | 0.019 |
| 3000 | 0.037 | 0.027 | 0.022 | 0.019 | 0.017 | 0.016 |
| 4000 | 0.028 | 0.020 | 0.017 | 0.014 | 0.013 | 0.012 |
| 5000 | 0.022 | 0.016 | 0.013 | 0.011 | 0.010 | 0.009 |

TABLE 3. CRITICAL VALUES OF $b_{2,2}^{+}$

| $n$ $2\alpha \rightarrow$ | 0.01 | | 0.025 | | 0.05 | | 0.10 | |
|---|---|---|---|---|---|---|---|---|
| | lower | upper | lower | upper | lower | upper | lower | upper |
| 10 | 4.580 | 10.378 | 4.722 | 9.781 | 4.887 | 9.203 | 5.057 | 8.606 |
| 12 | 4.732 | 10.881 | 4.899 | 10.150 | 5.053 | 9.593 | 5.232 | 8.947 |
| 14 | 4.842 | 11.159 | 5.015 | 10.375 | 5.179 | 9.769 | 5.358 | 9.162 |
| 16 | 4.977 | 11.387 | 5.149 | 10.562 | 5.318 | 9.941 | 5.482 | 9.331 |
| 18 | 5.045 | 11.478 | 5.219 | 10.628 | 5.382 | 10.005 | 5.555 | 9.403 |
| 20 | 5.175 | 11.609 | 5.262 | 10.691 | 5.533 | 10.114 | 5.717 | 9.469 |
| 25 | 5.351 | 11.628 | 5.525 | 10.584 | 5.689 | 10.159 | 5.871 | 9.503 |
| 30 | 5.518 | 11.594 | 5.692 | 10.556 | 5.855 | 10.156 | 6.038 | 9.516 |
| 40 | 5.703 | 11.453 | 5.871 | 10.563 | 6.139 | 10.109 | 6.229 | 9.497 |
| 50 | 5.909 | 11.181 | 6.083 | 1.0372 | 6.239 | 9.987 | 6.403 | 9.453 |
| 60 | 6.015 | 10.994 | 6.189 | 10.250 | 6.335 | 9.889 | 6.505 | 9.401 |
| 70 | 6.139 | 10.753 | 6.290 | 10.106 | 6.437 | 9.781 | 6.602 | 9.356 |
| 80 | 6.223 | 10.537 | 6.372 | 9.981 | 6.539 | 9.694 | 6.683 | 9.309 |
| 90 | 6.332 | 10.325 | 6.475 | 9.885 | 6.622 | 9.688 | 6.749 | 9.256 |
| 100 | 6.389 | 10.188 | 6.521 | 9.806 | 6.665 | 9.556 | 6.793 | 9.210 |
| 150 | 6.615 | 10.253 | 6.749 | 9.475 | 6.858 | 9.300 | 6.972 | 9.027 |
| 200 | 6.761 | 9.506 | 6.889 | 9.269 | 6.979 | 9.141 | 7.083 | 8.919 |
| 300 | 6.949 | 9.219 | 7.052 | 9.031 | 7.142 | 8.916 | 7.245 | 8.766 |
| 400 | 7.079 | 9.061 | 7.171 | 8.917 | 7.252 | 8.787 | 7.342 | 8.664 |
| 600 | 7.232 | 8.874 | 7.295 | 8.749 | 7.369 | 8.647 | 7.464 | 8.547 |
| 800 | 7.304 | 8.747 | 7.372 | 8.641 | 7.451 | 8.562 | 7.536 | 8.472 |
| 1000 | 7.367 | 8.656 | 7.433 | 8.569 | 7.504 | 8.497 | 7.585 | 8.419 |
| 1500 | 7.460 | 8.532 | 7.537 | 8.463 | 7.595 | 8.405 | 7.661 | 8.339 |
| 2000 | 7.535 | 8.461 | 7.599 | 8.401 | 7.649 | 8.351 | 7.707 | 8.293 |
| 2500 | 7.588 | 8.412 | 7.641 | 8.359 | 7.686 | 8.314 | 7.738 | 8.262 |
| 3000 | 7.624 | 8.376 | 7.673 | 8.327 | 7.714 | 8.286 | 7.760 | 8.240 |
| 4000 | 7.674 | 8.326 | 7.716 | 8.284 | 7.752 | 8.248 | 7.793 | 8.207 |
| 5000 | 7.709 | 8.291 | 7.746 | 8.254 | 7.778 | 8.222 | 7.814 | 8.186 |

*Example :* We apply the above test of multivariate normality to Shewhart's data (1931, p. 42) of sixty measurements on tensile strength ($X$) and hardness ($Y$) of aluminium die-castings. Under the assumption of normality, Pearson and Wilks (1933) and various other authors have analysed this data. Table 4 gives the moments upto the fourth order of the data. The sample means are denoted by $\bar{X}$ and $\bar{Y}$, the sample variances by $S_1^2$ and $S_2^2$, the sample correlation by $r$, and $g_{rs} = M_{rs}/S_1^r S_2^s$, $M_{rs}$ being the $(r, s)$ central moment. On using the moments in Table 4, it is found that

$$b_{1,2} = 0.1322, \quad b_{2,2} = 6.4389.$$

TABLE 4. MOMENTS FOR SHEWHART'S DATA ON ALUMINIUM DIE CASTINGS; TENSILE STRENGTH ($X$) AND HARDNESS ($Y$), $n = 60$

| moments of $X$ | moments of $Y$ | moments of $(X, Y)$ |
|---|---|---|
| $\bar{X} = 31.8694$ | $\bar{Y} = 69.8250$ | $r = 0.6832$ |
| $S_1^2 = 15.7049$ | $S_2^2 = 138.6019$ | |
| $g_{30} = 0.0998$ | $g_{03} = 0.0070$ | $g_{12} = 0.0867$ |
| $g_{40} = 2.6056$ | $g_{04} = 2.1119$ | $g_{21} = 0.1587$ |
| | | $g_{13} = 1.6070$ |
| | | $g_{31} = 1.7945$ |
| | | $g_{22} = 1.6574$ |

The 5% value of $b_{1,2}$ from Table 2 is 0.906 whereas the lower and the upper 2.5% values of $b_{2,2}$ from Table 3 are 6.335 and 9.889 respectively. Hence, the data may be regarded as a sample from a bivariate normal population.

I wish to express my thanks to Dr. E. A. Evans for his helpful comments, to Professor R. A. Reyment for raising various practical queries, and to the referee for his valuable suggestions.

## References


Anderson, T. W. (1958) : *Introduction to Multivariate Statistical Analysis*, Wiley, New York.

Box, G. E. P. (1953) : Non-normality and tests on variances. *Biometrika* **43,** 318-35.

David, F. N., Kendall, M. G. and Barton, D. E. (1966) : *Symmetric Functions and Allied Tables*, Cambridge University Press.

Day, N. E. (1969) : Estimating the components of a mixture of normal distributions. *Biometrika,* **56,** 463-74.

Fisher, R. A. (1930): The moments of the distribution for normal samples of measures of departures from normality. *Proc. Roy. Soc.*, A, **130,** 16-28.

GEARY, R. C. (1933) : A general expression for the moments of certain symmetrical functions of normal samples. *Biometrika*, **25**, 184-86.

HOPKINS, J. W. and CLAY, P. P. F. (1963) : Some empirical distributions of bivariate $T^2$ and homoscedasticity, criterion $M$ under unequal variance and leptokurtosis. *J. Amer. Statist. Ass.*, **58**, 1048-53.

ITO, K. (1969) : On the effect of heteroscedasticity and nonnormality upon some multivariate test procedures. *Proc. Int. Symp. Multivariate Analysis*, **2**, 87-120 (ed. P. R. Krishnaiah). Academic Press, New York.

KENDALL, M. G. and STUART, A. (1967, 1969) : *The Advanced Theory of Statistics*, **2** (2nd ed.), **1** (3rd ed.). Griffin, London.

KHATRI, C. G. (1959) : On the mutual independence of certain statistics. *Ann. Math. Statist.* **30** 1258-62.

KHATRI C. G. and PILLAI K. C. S. (1967) : On the moments of traces of two matrices in multivariate analysis. *Ann. Inst. Statist. Math.* **19**, **143**-56.

MARDIA K. V. (1970) : Measures of multivariate skewness and kurtossis with applications. *Biometrika*, **57**, 519-530.

——— (1971) : The effect of nonnormality on some multivariate tests and robustness to nonnormality in the linear model. *Biometrika*, **58**, 105-21.

PEARSON, E. S. and HARTLEY H. O. (1966) : *Biometrika Tables for Statisticians*, **1**, (3rd edn.) Cambridge University Press.

PEARSON, E. S. and WILKS S. S. (1933) : Methods of statistical analysis appropriate for $k$ samples of two variables. *Biometrika*, **25**, 353-378.

RAO, C. R. (1965) : *Linear Statistical Inference and its Applications.*, Wiley, New York.

SHEWHART, W. A. (1931) : *Economic Control of Quality of Manufactured Product.*, Macmillan, London.

*Paper received : October, 1972.*
*Revised : July, 1973.*