Measures of Multivariate Skewness and Kurtosis with Applications

Author(s): K. V. Mardia

Source: *Biometrika*, Dec., 1970, Vol. 57, No. 3 (Dec., 1970), pp. 519-530
Published by: Oxford University Press on behalf of Biometrika Trust
Stable URL: https://www.jstor.org/stable/2334770

*Biometrika* (1970), 57, 3, *p.* 519
*Printed in Great Britain*



# Measures of multivariate skewness and kurtosis with applications

By K. V. MARDIA
*University of Hull*

Summary

Measures of multivariate skewness and kurtosis are developed by extending certain studies on robustness of the $t$ statistic. These measures are shown to possess desirable properties. The asymptotic distributions of the measures for samples from a multivariate normal population are derived and a test of multivariate normality is proposed. The effect of nonnormality on the size of the one-sample Hotelling's $T^2$ test is studied empirically with the help of these measures, and it is found that Hotelling's $T^2$ test is more sensitive to the measure of skewness than to the measure of kurtosis.

## 1. Introduction

A popular measure of univariate skewness is $\beta_1$ and of univariate kurtosis is $\beta_2$. These measures have proved useful (i) in selecting a member of a family such as from the Karl Pearson family, (ii) in developing a test of normality, and (iii) in investigating the robustness of the standard normal theory procedures. The role of the tests of normality in modern statistics has recently been summarized by Shapiro & Wilk (1965).

With these applications in mind for the multivariate situations, we propose measures of multivariate skewness and kurtosis. These measures of skewness and kurtosis are developed naturally by extending certain aspects of some robustness studies for the $t$ statistic which involve $\beta_1$ and $\beta_2$. It should be noted that measures of multivariate dispersion have been available for quite some time (Wilks, 1932, 1960; Hotelling, 1951).

We deal with the measure of skewness in § 2 and with the measure of kurtosis in § 3. In § 4 we give two important applications of these measures, namely, a test of multivariate normality and a study of the effect of nonnormality on the size of the one-sample Hotelling's $T^2$ test. Both of these problems have attracted attention recently. The first problem has been treated by Wagle (1968) and Day (1969) and the second by Arnold (1964), but our approach differs from theirs.

## 2. A measure of skewness

### 2·1. *Notation and intuitive development*

Let the mean and variance of a random sample of size $n$ from a nonnormal population be $\bar{X}$ and $S^2$, respectively. Let $\kappa_r$ denote the $r$th cumulant for the population and write $\gamma_2 = \beta_2 - 3$. Kendall & Stuart (1967, p. 466) have investigated the effect of nonnormality on the one-sample $t$ test with the help of the asymptotic result

$$\operatorname{corr}(\bar{X}, S^2) \sim \{\beta_1/(2+\gamma_2)\}^{\frac{1}{2}}. \tag{2·1}$$

33-2

The result (2·1) is obtained by taking $\operatorname{var}(\bar{X})$, $\operatorname{var}(S^2)$ and $\operatorname{cov}(\bar{X}, S^2)$ to order $n^{-1}$. If we further assume that $\kappa_4$ is negligible, then (2·1) reduces to

$$\operatorname{corr}(\bar{X}, S^2) \sim (\tfrac{1}{2}\beta_1)^{\frac{1}{2}}. \tag{2·2}$$

Consequently, under the above assumptions, $\operatorname{corr}(\bar{X}, S^2)$ can be regarded as a measure of univariate skewness.

This discussion suggests the following extension of $\beta_1$ to the multivariate case. Let $\mathbf{X}_i' = (X_{1i}, \ldots, X_{pi})$ $(i = 1, 2, \ldots, n)$ be a random sample of size $n$ from a $p$-variate population with random vector $\mathbf{X}' = (X_1, \ldots, X_p)$. Let $\bar{\mathbf{X}}' = (\bar{X}_1, \ldots, \bar{X}_p)$ and $\mathbf{S} = \{S_{ij}\}$ denote the sample mean vector and the covariance matrix respectively. Let $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ be the mean vector and the covariance matrix of $\mathbf{X}$. We assume in this section for simplification that $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma}$ is a nonsingular matrix. A suitable measure of multivariate skewness, denoted by $\beta_{1,p}$, may be obtained by considering the canonical correlations between $\bar{\mathbf{X}}$ and $\mathbf{S}$ under the following assumptions which are analogous to those used in the derivation of (2·2).

(i) The second order moments of $\bar{\mathbf{X}}$ and $\mathbf{S}$ are taken to order $n^{-1}$.

(ii) The cumulants of order higher than 3 of $\mathbf{X}$ are negligible. Let us write the elements of $\mathbf{S} = \{S_{ij}\}$ as the vector

$$\mathbf{u} = (S_{11}, \ldots, S_{pp}, S_{12}, \ldots, S_{1p}, \ldots, S_{2p}, \ldots, S_{p-1,p})',$$

which has $p+q$ elements, $q = \frac{1}{2}p(p-1)$. The first $p$ components of $\mathbf{u}$ are the diagonal elements of $\mathbf{S}$ while the remaining $q$ components are the elements above the main diagonal of $\mathbf{S}$. Let us denote the covariance matrix of the random vector $(\bar{\mathbf{X}}, \mathbf{u})$ by

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}_{11} & \boldsymbol{\Lambda}_{12} \\ \boldsymbol{\Lambda}_{21} & \boldsymbol{\Lambda}_{22} \end{bmatrix},$$

which means that $\boldsymbol{\Lambda}_{11}$ is the covariance matrix of $\bar{\mathbf{X}}$ and so on. The canonical correlations $\lambda_1, \ldots, \lambda_p$, of $\bar{\mathbf{X}}$ and $\mathbf{u}$ are the roots of the determinantal equation (Kendall & Stuart, 1968, p. 305)

$$|\boldsymbol{\Lambda}_{11}^{-1}\boldsymbol{\Lambda}_{12}\boldsymbol{\Lambda}_{22}^{-1}\boldsymbol{\Lambda}_{21} - \lambda^2\mathbf{I}| = 0, \tag{2·3}$$

where $\mathbf{I}$ is the identity matrix of order $p \times p$. We can take any function of $\lambda_1^2, \ldots, \lambda_p^2$ as a measure of skewness $\beta_{1,p}$ provided it satisfies a certain invariance property. The most natural choice of the function is

$$\beta_{1,p} = 2\sum_{i=1}^{p} \lambda_i^2. \tag{2·4}$$

The multiplier 2 is introduced so that the value of $\beta_{1,p}$ for $p = 1$ reduces to (2·1). It will be shown subsequently that $\beta_{1,p}$ defined by (2·4) possesses the invariance property.

We shall first simplify $\beta_{1,p}$ for $\boldsymbol{\Sigma} = \mathbf{I}$ and then extend $\beta_{1,p}$ for general $\boldsymbol{\Sigma}$.

### 2·2. *The measure $\beta_{1,p}$ for $\boldsymbol{\Sigma} = \mathbf{I}$*

From (2·3) and (2·4), we have

$$\beta_{1,p} = 2\operatorname{tr}(\boldsymbol{\Lambda}_{11}^{-1}\boldsymbol{\Lambda}_{12}\boldsymbol{\Lambda}_{22}^{-1}\boldsymbol{\Lambda}_{21}). \tag{2·5}$$

Since $\boldsymbol{\Lambda}_{11} = \mathbf{I}/n$, (2·5) reduces to

$$\beta_{1,p} = 2n\operatorname{tr}(\boldsymbol{\Lambda}_{22}^{-1}\boldsymbol{\Lambda}_{21}\boldsymbol{\Lambda}_{12}). \tag{2·6}$$

We have, to order $n^{-1}$,

$$\boldsymbol{\Lambda}_{22} = (1/n)\operatorname{diag}(2, \ldots, 2, 1, \ldots, 1)$$

with $p$ 2's and $q$ 1's and on substituting for $\mathbf{\Lambda}_{22}$ in (2·6), we have

$$\beta_{1,p} = n^2 \sum_{i,j,k=1}^{p} \{\operatorname{cov}(\bar{X}_i, S_{jk})\}^2. \tag{2·7}$$

Under the assumptions (i) and (ii) of § 2·1 and $\boldsymbol{\mu} = \mathbf{0}$, it can be shown that we have to order $n^{-1}$

$$\operatorname{cov}(\bar{X}_i, S_{jk}) = (1/n)\, E(X_i X_j X_k)$$

for all $i, j$ and $k$, so that (2·7) becomes

$$\beta_{1,p} = \sum_{i,j,k} \{E(X_i X_j X_k)\}^2. \tag{2·8}$$

2·3. *A general lemma for the extension*

The extension of (2·8) for $\mathbf{\Sigma}$ will be obtained with the help of a general lemma. The lemma will also be useful for the measure of multivariate kurtosis treated here.

Let $\Phi(\mathbf{X})$ be a functional defined on a random vector $\mathbf{X}$ with mean vector zero and covariance matrix $\mathbf{I}$. Let $\mathbf{\Sigma}$ be a positive definite matrix. Then there exists a nonsingular matrix $\mathbf{U}$ such that

$$\mathbf{U\Sigma U'} = \mathbf{I}. \tag{2·9}$$

This implies that the random vector $\mathbf{Y}$ defined by

$$\mathbf{X} = \mathbf{UY} \tag{2·10}$$

will obviously have the functional corresponding to $\Phi$ as

$$\Phi^*(\mathbf{Y}) \equiv \Phi(\mathbf{UY}). \tag{2·11}$$

In view of (2·9) the covariance matrix of $\mathbf{Y}$ is $\mathbf{\Sigma}$. However, (2·9) does not define $\mathbf{U}$ uniquely for a given $\mathbf{\Sigma}$ and so $\Phi^*$ defined by (2·11) may not extend $\Phi$ uniquely. The following lemma gives a sufficient condition under which $\Phi^*$ will be unique.

LEMMA 1. *The functional $\Phi^*(\mathbf{Y})$ is uniquely defined for a given $\mathbf{\Sigma}$ if $\Phi(\mathbf{X})$ is invariant under orthogonal transformations. Furthermore $\Phi^*(\mathbf{Y})$ will then be invariant under nonsingular transformations.*

*Proof.* Let $\mathbf{V}$ be another matrix satisfying (2·9), i.e.

$$\mathbf{V\Sigma V'} = \mathbf{I}. \tag{2·12}$$

For the first part of the lemma, it is sufficient to show that

$$\Phi(\mathbf{UY}) = \Phi(\mathbf{VY}). \tag{2·13}$$

From (2·9) and (2·12) we find that

$$\mathbf{V} = \mathbf{CU}, \tag{2·14}$$

where $\mathbf{C}$ is an orthogonal matrix. Under the assumption of the lemma, we have

$$\Phi(\mathbf{UY}) = \Phi(\mathbf{X}) = \Phi(\mathbf{CX}).$$

From the preceding equation and (2·10) we obtain $\Phi(\mathbf{UY}) = \Phi(\mathbf{CUY})$, which on using (2·14) leads to (2·13).

We now prove the invariance property. Let

$$\mathbf{Y}_1 = \mathbf{AY} \tag{2·15}$$

be a nonsingular transformation and let $\mathbf{\Sigma}_1$ denote the covariance matrix of $\mathbf{Y}_1$. We have

$$(\mathbf{UA}^{-1})\, \mathbf{\Sigma}_1 (\mathbf{UA}^{-1})' = \mathbf{I}.$$

Therefore, from (2·11),

$$\Phi^*(\mathbf{Y}_1) = \Phi(\mathbf{U}\mathbf{A}^{-1}\mathbf{Y}_1),$$

which on using (2·11) and (2·15) reduces to

$$\Phi^*(\mathbf{Y}_1) = \Phi(\mathbf{U}\mathbf{Y}) = \Phi^*(\mathbf{Y}).$$

Hence the lemma follows.

### 2·4. *The measure $\beta_{1,p}$ for general $\mathbf{\Sigma}$ and its properties*

We now verify that (2·8) satisfies the condition of Lemma 1, i.e. invariance under the orthogonal transformation $\mathbf{Y} = \mathbf{C}\mathbf{X}$. Let $\mathbf{C} = \{C_{ij}\}$. Then

$$X_i = \sum_{r=1}^{p} C_{ri} Y_r.$$

On substituting for $X_i$ in (2·8) we have, after some simplification,

$$\beta_{1,p} = \sum_{r,s,t} \sum_{r',s',t'} \left(\sum_{i=1}^{p} C_{ri} C_{r'i}\right)\left(\sum_{j=1}^{p} C_{sj} C_{s'j}\right)\left(\sum_{k=1}^{p} C_{rk} C_{r'k}\right) E(Y_r Y_s Y_t)\, E(Y_{r'} Y_{s'} Y_{t'}). \quad (2·16)$$

Since $\mathbf{C}$ is an orthogonal matrix, (2·16) becomes

$$\beta_{1,p} = \sum_{r,s,t} \{E(Y_r Y_s Y_t)\}^2.$$

Hence $\beta_{1,p}$ is invariant under orthogonal transformations.

We now obtain the extension of $\beta_{1,p}$ from §2·3. For a positive definite matrix $\mathbf{\Sigma}$, there exists an orthogonal matrix $\mathbf{C}$ such that

$$\mathbf{C}'\mathbf{\Sigma}\mathbf{C} = \mathbf{D}, \quad (2·17)$$

where $\mathbf{D} = \operatorname{diag}(d_1, \ldots, d_p)$ $(d_i > 0;\ i = 1, \ldots, p)$. Consequently, we may take $\mathbf{U} = \mathbf{D}^{-\frac{1}{2}}\mathbf{C}'$ in (2·9) and then, from (2·10), we find

$$X_i = \sum_{r=1}^{p} C_{ri} Y_r / d_i^{\frac{1}{2}}.$$

Let $\mathbf{\Sigma}^{-1} = \{\sigma^{rr'}\}$. On substituting for $X_i$ in (2·8) and using the result

$$\sigma^{rr'} = \sum_{i=1}^{p} C_{ri} C_{r'i} / d_i,$$

obtained from (2·17), we find after some simplification that $\beta_{1,p}$ for the random vector $\mathbf{Y}$ is given by

$$\beta_{1,p} = \sum_{r,s,t} \sum_{r',s',t'} \sigma^{rr'} \sigma^{ss'} \sigma^{tt'} E(Y_r Y_s Y_t)\, E(Y_{r'} Y_{s'} Y_{t'}). \quad (2·18)$$

Hence, for a random vector $\mathbf{X}$ with mean vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)$ and covariance matrix $\mathbf{\Sigma}$, we have

$$\beta_{1,p} = \sum_{r,s,t} \sum_{r',s',t'} \sigma^{rr'} \sigma^{ss'} \sigma^{tt'} \mu_{111}^{(rst)} \mu_{111}^{(r's't')}, \quad (2·19)$$

where $\mu_{111}^{(rst)} = E\{(X_r - \mu_r)(X_s - \mu_s)(X_t - \mu_t)\}$. From Lemma 1, the measure given by (2·19) will be invariant under the nonsingular transformation

$$\mathbf{X} = \mathbf{A}\mathbf{Y} + \mathbf{b}. \quad (2·20)$$

Hence, the measure at least satisfies the most desirable property.

Let us consider some particular cases of (2·19). For $p = 1$, we, of course, have $\beta_{1,1} = \beta_1$. For $p = 2$, let us denote

$$\sigma_1^2 = \operatorname{var}(X_1), \quad \sigma_2^2 = \operatorname{var}(X_2), \quad \rho = \operatorname{corr}(X_1, X_2), \quad \gamma_{rs} = \mu_{rs}/\sigma_1^r \sigma_2^s,$$

where $\mu_{r,s}$ is the central moment of order $(r, s)$ of $(X_1, X_2)$. From (2·19) we have

$$\beta_{1,2} = (1-\rho^2)^{-3}[\gamma_{30}^2+\gamma_{03}^2+3(1+2\rho^2)(\gamma_{12}^2+\gamma_{21}^2) \\ -2\rho^3\gamma_{30}\gamma_{03}+6\rho\{\gamma_{30}(\rho\gamma_{12}-\gamma_{21})+\gamma_{03}(\rho\gamma_{21}-\gamma_{12})-(2+\rho^2)\gamma_{12}\gamma_{21}\}]. \qquad (2\cdot21)$$

For $\rho = 0$ and $\sigma_1 = \sigma_2 = 1$ we have

$$\beta_{1,2} = \mu_{30}^2+\mu_{03}^2+3\mu_{12}^2+3\mu_{21}^2. \qquad (2\cdot22)$$

Hence, $\beta_{1,2}$ is zero for uncorrelated random variables if and only if $\mu_{30}$, $\mu_{03}$, $\mu_{12}$ and $\mu_{21}$ all vanish simultaneously. This is obviously a desirable property for any sensible measure of skewness. For any symmetric distribution about $\boldsymbol{\mu}$, we have $\beta_{1,p} = 0$. In particular, for a normal population $\beta_{1,p} = 0$.

For a random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$, the measure of skewness corresponding to $\beta_{1,p}$ is given by

$$b_{1,p} = \sum_{r,s,t}\sum_{r',s',t'} S^{rr'}S^{ss'}S^{tt'}M_{111}^{(rst)}M_{111}^{(r's't')}, \qquad (2\cdot23)$$

where

$$\mathbf{S}^{-1} = \{S^{ij}\} \quad \text{and} \quad M_{111}^{(rst)} = \frac{1}{n}\sum_{i=1}^{n}(X_{ri}-\bar{X}_r)(X_{si}-\bar{X}_s)(X_{ti}-\bar{X}_t).$$

It can be seen that (2·23) is invariant under the nonsingular transformation (2·20).

### 2·5. *The asymptotic distribution of $b_{1,p}$ and a test of multivariate skewness*

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be a random sample from $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Since $b_{1,p}$ is invariant under linear transformations, we assume $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}$. Now, $\mathbf{S}$ converges to $\boldsymbol{\Sigma}$ in probability. Therefore, from (2·23) we get

$$b_{1,p} \to \sum_{r,s,t}\{M_{111}^{(rst)}\}^2, \qquad (2\cdot24)$$

in probability. On writing $M_{111}^{(rrr)} = M_3^{(r)}$ and $M_{111}^{(rss)} = M_{12}^{(rs)}$ $(r \neq s)$ in (2·24), we have

$$b_{1,p} \to \{M_3^{(1)}\}^2+\ldots+3\{M_{21}^{(12)}\}^2+\ldots+6\{M_{111}^{(123)}\}^2+\ldots \qquad (2\cdot25)$$

in probability. It can be shown to order $n^{-1}$ that

$$E\{M_{111}^{(rst)}\} = 0, \quad \operatorname{var}\{M_3^{(1)}\} = 6/n, \quad \operatorname{var}\{M_{21}^{(12)}\} = 2/n, \quad \operatorname{var}\{M_{111}^{(123)}\} = 1/n,$$

$$\operatorname{cov}(M_{111}^{(rst)}, M_{111}^{(r's't')}) = 0 \quad \{(r,s,t) \neq (r',s',t')\}.$$

Using the normality of the vector $\{M_3^{(1)}, \ldots, M_{21}^{(12)}, \ldots, M_{111}^{(123)}, \ldots\}$ together with its mean vector and covariance matrix obtained from the preceding results, we find with the help of the well-known results on the limiting distributions of quadratic forms that

$$n[\{M_3^{(1)}\}^2+\ldots+3\{M_{21}^{(12)}\}^2+\ldots+6\{M_{111}^{(123)}\}^2+\ldots]/6$$

has a $\chi^2$ distribution with $p(p+1)(p+2)/6$ degrees of freedom. Hence, from (2·24),

$$A = nb_{1,p}/6 \qquad (2\cdot26)$$

has a $\chi^2$ distribution with $p(p+1)(p+2)/6$ degrees of freedom. For $p > 7$, we can use the approximation that $(2A)^{\frac{1}{2}}$ is $N[\{p(p+1)(p+2)-3\}/3, 1]$.

To test $\beta_{1,p} = 0$ for large samples, we calculate $A$ and reject the hypothesis for large values of $A$. Using the terminology of the univariate case, the rejection of $\beta_{1,p} = 0$ can be described as an indication of skewness in the distribution of $\mathbf{X}$.

## 3. A measure of kurtosis

### 3·1. *A measure with motivation*

For the one-sample Pitman's permutation test in the univariate case, Box & Andersen (1955) have shown that the square of the $t$ statistic for this case has approximately an $F$ distribution with $\delta$ and $\delta(n-1)$ degrees of freedom, where

$$\delta = 1+\frac{\beta_2-3}{n}+o\left(\frac{1}{n}\right). \tag{3·1}$$

Thus the coefficient of $1/n$ in $\delta$ provides a measure of univariate kurtosis, namely, $\beta_2$. We now show that the multivariate extension of this problem by Arnold (1964) gives a sensible measure of kurtosis $\beta_{2,p}$.

We denote as in §2 that $\mathbf{X}_1, \ldots, \mathbf{X}_n$ is a random sample from a $p$-variate population with random vector $\mathbf{X}$ having mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Let $T^2$ denote Hotelling's $T^2$ statistic for the one-sample problem. Using the permutation moments of $T^2$ given in Arnold (1964) and following the method of Box & Andersen (1955), we find that the distribution of

$$(n-p)\,T^2/\{p(n-1)\} \tag{3·2}$$

is approximately an $F$ distribution with $\delta p$ and $\delta(n-p)$ degrees of freedom, where

$$\delta = 1+\frac{E(b^*_{2,p})-(p+2)}{n[1-\{E(b^*_{2,p})/(n+2)\}]}, \tag{3·3}$$

with

$$b^*_{2,p} = \{(n+2)/n^2p\}\sum_{i=1}^{n}\{(\mathbf{X}_i-\boldsymbol{\mu})'\bar{\mathbf{S}}^{-1}(\mathbf{X}_i-\boldsymbol{\mu})\}^2.$$

Further, $\bar{\mathbf{S}}$ is the sample covariance matrix about $\boldsymbol{\mu}$. We can write (3·3) as

$$\delta = 1+(1/n)\,[\{\beta_{2,p}-p(p+2)\}/p]+o(1/n), \tag{3·4}$$

where

$$\beta_{2,p} = E\{(\mathbf{X}-\boldsymbol{\mu})'\,\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})\}^2. \tag{3·5}$$

On comparing (3·4) with (3·1), we may take $\beta_{2,p}$ as a measure of multivariate kurtosis. A similar measure has appeared in a robustness study by Box & Watson (1962).

### 3·2. *Properties of $\beta_{2,p}$*

We now show that $\beta_{2,p}$ has some desirable properties. For $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}$, (3·5) reduces to

$$\beta_{2,p} = E\{(\mathbf{X}'\mathbf{X})^2\}, \tag{3·6}$$

which is invariant under orthogonal transformations. Hence, from Lemma 1 of §2·3, $\beta_{2,p}$ defined by (3·5) is invariant under nonsingular transformations. This property can also be established directly from (3·5).

Let $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}$. On considering

$$E\left(a_0+a_1\sum_{i=1}^{p}X_i+a_2\sum_{i=1}^{p}X_i^2\right)^2 \geqslant 0$$

for real values of $a_0$, $a_1$ and $a_2$, we obtain

$$\beta_{2,p} \geqslant p^2+A^2/p, \tag{3·7}$$

where

$$A = E\left\{\left(\sum_{i=1}^{p} X_i\right)\left(\sum_{i=1}^{p} X_i^2\right)\right\}.$$

This implies that $\beta_{2,p} \geqslant p^2$. For $p = 1$, (3·7) reduces to the well-known inequality $\beta_2 \geqslant 1+\beta_1^2$ but $\beta_{1,p}$ and $\beta_{2,p}$ are not related in such a way.

Pearson (1919) has given a Tchebycheff-type inequality for the bivariate case. His argument is applicable to the $p$-variate case and we have for any $\epsilon > 0$

$$\mathrm{pr}\,\{(\mathbf{X}-\boldsymbol{\mu})'\,\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu}) < \epsilon\} \geqslant 1-\beta_{2,p}/\epsilon^2.$$

In view of these properties, $\beta_{2,p}$ may be regarded as a satisfactory measure of multivariate kurtosis.

The expression (3·5) can be simplified. In fact,

$$\beta_{2,p} = \Sigma(\sigma^{ii})^2\,\boldsymbol{\mu}_4^{(i)} + 4\Sigma\sigma^{ii}\sigma^{ij}\mu_{31}^{(ij)} + \Sigma\{2(\sigma^{ij})^2+\sigma^{ii}\sigma^{jj}\}\,\mu_{22}^{(ij)} + \Sigma\{4\sigma^{ij}\sigma^{jk}+2\sigma^{ii}\sigma^{jk}\}\,\mu_{211}^{(ijk)} + \Sigma\sigma^{ij}\sigma^{kl}\,\mu_{1111}^{(ijkl)}, \quad (3\cdot 8)$$

where

$$\mu_{i_1\ldots i_s}^{(j_1\ldots j_s)} = E\left\{\prod_{r=1}^{s}(X_{j_r}-\mu_{j_r})^{i_r}\right\}, \quad \boldsymbol{\Sigma}^{-1} = (\sigma^{ij})$$

and $\Sigma$ stands for the sum over all possible unequal values of the subscripts. For $p = 2$ we find that

$$\beta_{2,2} = \{\gamma_{40}+\gamma_{04}+2\gamma_{22}+4\rho(\rho\gamma_{22}-\gamma_{13}-\gamma_{31})\}/(1-\rho^2)^2, \quad (3\cdot 9)$$

where the notation of (2·21) is used. Furthermore, for $\sigma_1 = \sigma_2 = 1$ and $\rho = 0$,

$$\beta_{2,2} = \mu_{04}+\mu_{40}+2\mu_{22}. \quad (3\cdot 10)$$

For a normal population, we have

$$\beta_{2,p} = p(p+2). \quad (3\cdot 11)$$

For a random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$, the measure of kurtosis corresponding to $\beta_{2,p}$ is obviously

$$b_{2,p} = \frac{1}{n}\sum_{i=1}^{n}\{(\mathbf{X}_i-\overline{\mathbf{X}})'\,\mathbf{S}^{-1}(\mathbf{X}_i-\overline{\mathbf{X}})\}^2. \quad (3\cdot 12)$$

This is also invariant under the nonsingular transformation given by (2·20).

### 3·3. *The asymptotic distribution of $b_{2,p}$ and a test of multivariate kurtosis*

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be a random sample from $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. In view of the invariance property of $b_{2,p}$, we again assume $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}$. We first obtain the exact value of $E(b_{2,p})$.

#### 3·3·1. *The expectation of $b_{2,p}$*

From (3·12) we have

$$E(b_{2,p}) = E\{(\mathbf{X}_n-\overline{\mathbf{X}})'\,\mathbf{S}^{-1}(\mathbf{X}_n-\overline{\mathbf{X}})\}^2. \quad (3\cdot 13)$$

Let $\mathbf{X}_r^* = (X_{r1}, \ldots, X_{rn})'$ $(r = 1, \ldots, p)$. We transform $\mathbf{X}_r^*$ to $\boldsymbol{\xi}_r^* = (\xi_{r1}, \ldots, \xi_{rn})'$ by an orthogonal transformation $\boldsymbol{\xi}_r^* = \mathbf{C}\mathbf{X}_r^*$ $(r = 1, \ldots, p)$, where $\mathbf{C}$ is an orthogonal matrix with the first row as $(n^{-\frac{1}{2}}, \ldots, n^{-\frac{1}{2}})$ and the second row as $(-a, \ldots, -a, 1/na)$, $a$ being $\{n(n-1)\}^{-\frac{1}{2}}$. It is found that (3·13) reduces to

$$E(b_{2,p}) = (n-1)^2 E(Y^2), \quad (3\cdot 14)$$

where

$$Y = \boldsymbol{\xi}_2'\left(\sum_{t=2}^{n}\boldsymbol{\xi}_t\boldsymbol{\xi}_t'\right)^{-1}\boldsymbol{\xi}_2 \quad \text{with} \quad \boldsymbol{\xi}_t' = (\xi_{1t}, \ldots, \xi_{pt}) \quad (t = 2, \ldots, n).$$

Furthermore, $\boldsymbol{\xi}_t'$ $(t = 2, \ldots, n)$ are distributed independently as $N(\mathbf{0}, \mathbf{I})$. Let

$$\mathbf{S}_1 = \sum_{t=3}^{n} \boldsymbol{\xi}_t \boldsymbol{\xi}_t'.$$

We have

$$1 - Y = |\mathbf{S}_1|/|\mathbf{S}_1 + \boldsymbol{\xi}_2 \boldsymbol{\xi}_2'|$$

so that, from Rao (1965, p. 459), the probability density function of $Y$ is

$$\left\{\mathrm{B}\left(\frac{p}{2}, \frac{n-p-1}{2}\right)\right\}^{-1} y^{(\frac{1}{2}p)-1}(1-y)^{\frac{1}{2}(n-p-1)-1} \quad (0 < y < 1).$$

Hence,

$$E(Y^2) = p(p+2)/\{(n+1)(n-1)\}, \tag{3·15}$$

and on substituting (3·15) in (3·14) we finally have

$$E(b_{2,p}) = p(p+2)(n-1)/(n+1). \tag{3·16}$$

3·3·2. *The asymptotic variance of* $b_{2,p}$

We have succeeded in obtaining the exact value of $E(b_{2,p})$. By the same method, $E(b_{2,p}^2)$ can be simplified if $E(YZ)^2$ can be obtained explicitly, where

$$Z = \left\{\boldsymbol{\xi}_3' \left(\sum_{t=2}^{n} \boldsymbol{\xi}_t \boldsymbol{\xi}_t'\right)^{-1} \boldsymbol{\xi}_3\right\}.$$

However, it seems $E(YZ)^2$ cannot be simplified. The joint distribution of $Y$ and $Z$ is a bivariate beta but is not a particular case of various known multivariate beta distributions; see Olkin & Rubin (1964).

We now proceed to obtain the value of var $(b_{2,p})$ to order $n^{-1}$ on following a method used by Lawley (1959) for a different problem. Let $\mathbf{S} = \mathbf{I} + \mathbf{S}^*$ so that to order $n^{-1}$, $E(\mathbf{S}^*) = \mathbf{0}$. On using

$$\mathbf{S}^{-1} = (\mathbf{I} + \mathbf{S}^*)^{-1} = \mathbf{I} - \mathbf{S}^* + \mathbf{S}^{*2} - \ldots$$

in (3·12), we obtain after some simplification that

$$b_{2,p} = \frac{1}{n}\sum_{i=1}^{n} \{(\mathbf{X}_i - \overline{\mathbf{X}})'(\mathbf{X}_i - \overline{\mathbf{X}})\}^2 - \frac{2}{n}\sum_{i=1}^{n} (\mathbf{X}_i - \overline{\mathbf{X}})'(\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i - \overline{\mathbf{X}})'\mathbf{S}^*(\mathbf{X}_i - \overline{\mathbf{X}}) + \ldots. \tag{3·17}$$

It can be shown from the results of Cramér (1946, pp. 346–9) that the remaining terms in (3·17) will contribute quantities of orders lower than $n^{-1}$ in var $(b_{2,p})$. We can rewrite (3·17) as

$$b_{2,p} = \sum_{i=1}^{p} M_4^{(i)} + \sum_{i \neq j} M_{22}^{(ij)} - 2\sum_{i=1}^{p} M_2^{(i)*} M_4^{(i)} - 2\sum_{i \neq j} M_2^{(j)*} M_{22}^{(ij)} - 2\sum_{i=1}^{p}\sum_{j \neq k} M_{11}^{(jk)*} M_{211}^{(ijk)} + \ldots, \tag{3·18}$$

where

$$M_{i_1 \ldots i_s}^{(j_1 \ldots j_s)} = \frac{1}{n}\sum_{i=1}^{n}\left\{\prod_{r=1}^{s} (X_{j_r i} - \overline{X}_{j_r})^{i_r}\right\}, \quad M_2^{(i)*} = S_{ii}^*, \quad M_{11}^{(ij)*} = S_{ij}^*$$

with $\mathbf{S}^* = \{S_{ij}^*\}$. Applying the asymptotic formula for calculating sampling variances of Kendall & Stuart (1969, pp. 231–2) to $b_{2,p}$ given by (3·18) and using the asymptotic results,

$$E(\mathbf{S}^*) = \mathbf{0}, \quad \mathrm{var}\{M_2^{(i)}\} = 2/n, \quad \mathrm{var}\{M_4^{(i)}\} = 96/n, \quad \mathrm{var}\{M_{22}^{(ij)}\} = 8/n,$$

$$\mathrm{cov}\{M_2^{(i)}, M_2^{(j)}\} = \mathrm{cov}\{M_4^{(i)}, M_4^{(j)}\} = \mathrm{cov}\{M_4^{(i)}, M_{22}^{(jk)}\} = 0 \quad (i \neq j \neq k),$$

$$\mathrm{cov}\{M_2^{(i)}, M_4^{(i)}\} = 12/n, \quad \mathrm{cov}\{M_2^{(i)}, M_{22}^{(ij)}\} = 2/n,$$

$$\mathrm{cov}\{M_4^{(i)}, M_{22}^{(ij)}\} = 12/n, \quad \mathrm{cov}\{M_{22}^{(ij)}, M_{22}^{(ik)}\} = 2/n \quad (i \neq j \neq k),$$

we find after some algebra that to order $n^{-1}$

$$\operatorname{var}(b_{2,p}) = 8p(p+2)/n. \tag{3·19}$$

3·3·2. *The asymptotic distribution of* $b_{2,p}$

We have

$$b_{2,p} \to \frac{1}{n}\sum_{i=1}^{n}\{(\mathbf{X}_i-\overline{\mathbf{X}})'\,\mathbf{\Sigma}^{-1}(\mathbf{X}_i-\overline{\mathbf{X}})\}^2$$

in probability. On using the results given by (3·16) and (3·19) and the central limit theorem we find that

$$[b_{2,p}-\{p(p+2)(n-1)/(n+1)\}]/\{8p(p+2)/n\}^{\frac{1}{2}} \tag{3·20}$$

is asymptotically distributed as $N(0,1)$. Hence, we can test $\beta_{2,p} = p(p+2)$ for large samples. The rejection of the null hypothesis can be described as an indication of kurtosis in the probability density function of $\mathbf{X}$.

It should be noted that the asymptotic distribution of $nb_{1,p}/6$ derived in § 2·5 is $\chi^2$ rather than normal. This seems intuitively inevitable for any sensible measures of skewness and kurtosis. For example, in the bivariate case with $\mathbf{\Sigma} = \mathbf{I}$, a sensible measure of skewness must accumulate the effects of $\mu_{21}$, $\mu_{12}$, $\mu_{30}$ and $\mu_{03}$, which in $\beta_{1,2}$, given by (2·22), is achieved by taking a weighted sum of squares of these moments. However, $\mu_{22}$, $\mu_{40}$ and $\mu_{04}$ are positive so that a weighted sum of these moments as appearing in $\beta_{2,2}$ given by (3·10) is sufficient to measure the kurtosis. These basic differences will be reflected in the distributions of the corresponding sample characteristics.

## 4. Applications

### 4·1. *A test of multivariate normality*

It can be shown for samples from a normal population that the covariance matrix of $\{M_3^{(1)}, \ldots, M_{21}^{(12)}, \ldots, M_{111}^{(123)}, \ldots\}$ and $\{M_2^{(1)}, \ldots, M_4^{(1)}, \ldots, M_{22}^{(12)}, \ldots\}$ to order $n^{-1}$ is zero where the $M$'s are defined in §§ 2·5 and 3·3·2. Therefore, for large samples, we can test the normality by testing $\beta_{1,p} = 0$ and $\beta_{2,p} = p(p+2)$ separately. To test $\beta_{1,p} = 0$, we can use the result, from § 2·5, that

$$A = \tfrac{1}{6}nb_{1,p} \tag{4·1}$$

has a $\chi^2$ distribution with $p(p+1)(p+2)/6$ degrees of freedom and to test $\beta_{2,p} = p(p+2)$, we can use the result, from § 3·3·2, that

$$B = (b_{2,p}-\beta_{2,p})/\{8p(p+2)/n\}^{\frac{1}{2}} \tag{4·2}$$

is distributed as $N(0,1)$. Since $n$ is large, $B$ is asymptotically equivalent to (3·20).

To see the effectiveness of the test of normality, we applied it to the two sets of data described below.

*Data* 1. K. Pearson and A. Lee's data of the heights of 1376 fathers and daughters which is easily accessible (Fisher, 1963, pp. 178–9).

*Data* 2. Norton's data of the rate of discount and the ratio of reserves to deposits in 780 American Banks quoted by Yule & Kendall (1949, p. 201).

Tables 1 and 2 give the moments with Sheppard's correction for data 1 and data 2 respectively. The sample means are denoted by $\overline{X}$ and $\overline{Y}$, the sample variances by $S_1^2$ and $S_2^2$, the sample correlation coefficient by $r$, and $g_{rs} = M_{rs}/S_1^r S_2^s$, where $M_{rs}$ is the $(r, s)$ central moment for the sample.

On using the moments of Table 1, in the results corresponding to (2·21) and (3·9) for samples, we find that for data 1 $b_{1,2} = 0{\cdot}0335$ and $b_{2,2} = 8{\cdot}2881$. From (4·1) we have $A = 7{\cdot}68$. The 5 % value of $\chi_4^2$ is 9·49. Therefore $b_{1,2}$ is not significant. Further, using (4·2), we find $B = 1{\cdot}34$ which is less than 1·96, so that the difference $b_{2,2} - 8$ is not significant. Hence, we conclude that data 1 may be regarded as a sample from a normal population. Lancaster (1958) arrived at the same conclusion by another test.

Table 1. *Moments for data* 1 *of the heights of fathers* ($X$) *and daughters* ($Y$), $n = 1376$

| Moments of $X$ | Moments of $Y$ | Moments of $(X, Y)$ |
|---|---|---|
| $\bar{X} = 67{\cdot}6893$ | $\bar{Y} = 63{\cdot}8492$ | $r = 0{\cdot}5157$ |
| $S_1^2 = 7{\cdot}5527$ | $S_2^2 = 6{\cdot}6926$ | |
| | | $g_{12} = -0{\cdot}0009$ |
| $g_{30} = -0{\cdot}1545$ | $g_{03} = -0{\cdot}0078$ | $g_{21} = -0{\cdot}0328$ |
| $g_{40} = 2{\cdot}9731$ | $g_{04} = 3{\cdot}1647$ | $g_{13} = 1{\cdot}5294$ |
| | | $g_{31} = 1{\cdot}4853$ |
| | | $g_{22} = 1{\cdot}4735$ |

Table 2. *Moments for data* 2 *of the rate of discount* ($X$) *and the ratio of reserves to desposits* ($Y$), $n = 780$

| Moments of $X$ | Moments of $Y$ | Moments of $(X, Y)$ |
|---|---|---|
| $\bar{X} = 3{\cdot}3039$ | $\bar{Y} = 29{\cdot}8115$ | $r = -0{\cdot}5375$ |
| $S_1^2 = 7{\cdot}0006$ | $S_2^2 = 17{\cdot}8811$ | |
| | | $g_{12} = -0{\cdot}2940$ |
| $g_{30} = 3{\cdot}8055$ | $g_{03} = 1{\cdot}2887$ | $g_{21} = -0{\cdot}4714$ |
| $g_{40} = 27{\cdot}7573$ | $g_{04} = 4{\cdot}2135$ | $g_{13} = -1{\cdot}5891$ |
| | | $g_{31} = -3{\cdot}6416$ |
| | | $g_{22} = 1{\cdot}1857$ |

Similarly, on using the moments given in Table 2, we find that for data 2, $b_{1,2} = 28{\cdot}2306$ and $b_{2,2} = 48{\cdot}3873$, which give $A = 3669{\cdot}98$ and $B = 141{\cdot}03$. Both of these values are highly significant and therefore data 2 cannot be regarded as a sample from a normal population. In fact, the bivariate histogram drawn for this data by Yule & Kendall (1949, p. 204 A) indicates that the data are very skew.

### 4·2. *The effect of nonnormality on Hotelling's* $T^2$

Various theoretical and empirical studies have established that the size of the one-sample $t$ test is more sensitive to $\beta_1$ than to $\beta_2$; see, for example, the theoretical study by Bartlett (1935) and the recent empirical study by Ratcliffe (1968). In fact, Kendall & Stuart (1967, p. 466) have demonstrated this result ingeniously with the help of (2·1). By analogy, for sensible measures $\beta_{1,p}$ and $\beta_{2,p}$ it may be expected that the size of Hotelling's $T^2$ in the one-sample case will be influenced more by $\beta_{1,p}$ than by $\beta_{2,p}$. We investigate this problem by Monte Carlo methods.

Arnold (1964) has shown by Monte Carlo methods that the size of the one-sample Hotelling's $T^2$ test is not much influenced if samples are drawn from bivariate distributions of independent random variables when both marginal distributions are either of rectangular or of double exponential form. In fact, using (2·22) and (3·10), we can see that $\beta_{1,2} = 0$ for both bivariate distributions while $\beta_{2,2} = 5{\cdot}6$ for the bivariate rectangular distribution and $\beta_{2,2} = 14$ for the bivariate double exponential distribution. It should be noted that for any bivariate normal distribution, $\beta_{1,2} = 0$ and $\beta_{2,2} = 8$. Hence, the study suggests that $\beta_{2,2}$ does not have much effect on $T^2$. On the other hand, to see whether $\beta_{1,2}$ can have an appreciable effect on $T^2$, we consider samples from a bivariate distribution of independent random variables when both marginal distributions are of negative exponential form. For this distribution, we have $\beta_{1,2} = 8$ and $\beta_{2,2} = 20$. A pseudo-random number generator (Behrenz,

Table 3. *Actual* $100\alpha\%$ *significance level of Hotelling's* $T^2$ *test for samples from normal, rectangular and negative exponential populations*

| | | Exponential population $\beta_{1,2} = 8, \beta_{2,2} = 20$ (actual %) | | Normal population $\beta_{1,2} = 0, \beta_{2,2} = 8$ (actual %) | | Rectangular population* $\beta_{1,2} = 0, \beta_{2,2} = 5·6$ (actual %) | |
|---|---|---|---|---|---|---|---|
| $n$ | Nominal (%) | Estimate | 95 % confidence limits | Estimate | 95 % confidence limits | Estimate | 95 % confidence limits |
| 4 | 5·0 | 13·14 | 12·48–13·80 | 4·91 | 4·49–5·33 | 7·02 | 6·45–7·59 |
| | 2·5 | 8·54 | 7·99–9·09 | 2·40 | 2·10–1·70 | 3·63 | 3·18–4·08 |
| | 1·0 | 3·63 | 3·26–4·00 | 0·98 | 0·79–1·17 | 1·61 | 1·31–1·91 |
| | 0·5 | 2·15 | 1·87–2·43 | 0·54 | 0·40–0·68 | 0·88 | 0·65–1·11 |
| 6 | 5·0 | 14·81 | 14·11–15·51 | 5·32 | 4·88–5·76 | 6·44 | 6·04–6·84 |
| | 2·5 | 10·31 | 9·71–10·91 | 2·43 | 2·13–2·73 | 3·70 | 3·35–4·05 |
| | 1·0 | 5·92 | 5·46–6·38 | 0·87 | 0·69–1·05 | 1·83 | 1·53–2·13 |
| | 0·5 | 3·66 | 3·29–4·03 | 0·46 | 0·33–0·59 | 1·02 | 0·79–1·25 |
| 8 | 5·0 | 15·12 | 14·42–15·82 | 5·20 | 4·76–5·64 | 6·02 | 5·58–6·46 |
| | 2·5 | 10·51 | 9·91–11·11 | 2·59 | 2·28–2·90 | 3·24 | 2·94–3·54 |
| | 1·0 | 6·51 | 6·03–6·99 | 1·03 | 0·83–1·23 | 1·68 | 1·42–1·94 |
| | 0·5 | 4·73 | 4·31–5·15 | 0·47 | 0·34–0·60 | 0·96 | 0·78–1·14 |

* Values taken from Table 1 of Arnold (1964).

1962) was used for drawing samples of sizes $n = 4$, 6 and 8, from this population. The number of trials in each case was 10,000. We obtained the number of values of $T^2$ which exceeded $100\alpha\%$ values of the normal theory $T^2$ for $\alpha = 0·05$, 0·025, 0·01 and 0·005. The results are shown in Table 3 together with the 95% confidence limits obtained by using $\hat{p} \pm 1·96\{\hat{p}(1-\hat{p})/10{,}000\}^{\frac{1}{2}}$, where $\hat{p}$ is the actual level from the empirical trials. The results of Arnold (1964) for samples from the bivariate rectangular distribution are also shown in Table 3 for comparison.

We know that the pseudo-random numbers can introduce serious bias into the realizations of the problem under study; see Page (1965). A check on the suitability of the generator for our problem, as well as a check on the whole experimental process, was first obtained by taking the parent population normal. The results for this case are also shown in Table 3. The method of Box & Muller (1958) was used to generate the normal pairs.

It can be seen from Table 3 that the nominal levels for samples from the normal population agree well with the actual levels. However, the deviations from the nominal levels are very large for samples from the negative exponential population. Hence, this study indicates that Hotelling's $T^2$ test is more sensitive to $\beta_{1,p}$ than $\beta_{2,p}$. Furthermore, the actual level of significance of Hotelling's $T^2$ test when used on small samples from extremely nonnormal populations can be as much as three times the nominal level of 5 %.

I wish to express my thanks to Dr E. A. Evans and Dr J. W. Thompson for their helpful comments and to Mr M. J. Norman for his assistance in programming.

## References


Arnold, H. J. (1964). Permutation support for multivariate techniques. *Biometrika* **51**, 65–70.

Bartlett, M. S. (1935). The effect of non-normality on the $t$ distribution. *Proc. Camb. Phil. Soc.* **31**, 223–31.

Behrenz, P. G. (1962). Random number generator. *Comm. Ass. Computing Machinery* **5**, 553.

Box, G. E. P. & Andersen, S. L. (1955). Permutation theory in the derivation of robust criteria and the study of departures from assumptions. *J. R. Statist. Soc.* B **17**, 1–34.

Box, G. E. P. & Muller, M. E. (1958). A note on the generation of normal deviates. *Ann. Math. Statist.* **29**, 610–11.

Box, G. E. P. & Watson, G. S. (1962). Robustness to non-normality of regression tests. *Biometrika* **49**, 93–106. [Correction **52**, 669.]

Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press.

Day, N. E. (1969). Divisive cluster analysis and a test for multivariate normality. *Bull. Inst. Int. Statist.* **43**, II, 110–12.

Fisher, R. A. (1963). *Statistical Methods for Research Workers*, 13th ed. Edinburgh: Oliver and Boyd.

Hotelling, H. (1951). A generalized $T$ test and measure of multivariate dispersion. *Proc. Second Berkeley Symp.* pp. 23–42.

Kendall, M. G. & Stuart, A. (1967). *The Advanced Theory of Statistics*, vol. 2, 2nd edition. London: Griffin.

Kendall, M. G. & Stuart, A. (1968). *The Advanced Theory of Statistics*, vol. 3, 2nd edition. London: Griffin.

Kendall, M. G. & Stuart, A. (1969). *The Advanced Theory of Statistics*, vol. 1, 3rd edition. London: Griffin.

Lancaster, H. O. (1958). The structure of bivariate distributions. *Ann. Math. Statist.* **29**, 719–36.

Lawley, D. N. (1959). Tests of significance in canonical analysis. *Biometrika* **46**, 59–66.

Olkin, I. & Rubin, H. (1964). Multivariate beta distributions and independence properties of the Wishart distributions. *Ann. Math. Statist.* **35**, 261–9.

Page, E. S. (1965). The generation of pseudo-random numbers. In *Digital Simulation in Operational Research* (ed. S. H. Hollingdale), pp. 55–63. London: English Universities Press.

Pearson, K. (1919). On generalized Tchebycheff theorems in mathematical theory of statistics. *Biometrika* **12**, 284–96.

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

Ratcliffe, J. F. (1968). The effect on the $t$ distribution of non-normality in the sampled population. *Appl. Stat.* **17**, 42–8.

Shapiro, S. S. & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika* **52**, 591–611.

Wagle, B. (1968). Multivariate beta distribution and a test for multivariate normality. *J. R. Statist. Soc.* B **30**, 511–16.

Wilks, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika* **24**, 471–94.

Wilks, S. S. (1960). Multidimensional statistical scatter. In *Contributions to Probability and Statistics in Honor of Harold Hotelling*, ed. I. Olkin *et al.* Stanford University Press. pp. 486–503.

Yule, G. U. & Kendall, M. G. (1949). *An Introduction to the Theory of Statistics*, 13th ed. London: Griffin.


[*Received December* 1969. *Revised March* 1970]