MULTIVARIATE DATA ANALYSIS

S. B. Provost
Statistical and Actuarial Sciences, The University of Western Ontario, London, Ontario, Canada

E. M. Rudiuk
Physical and Applied Sciences, University College of Cape Breton, Sydney, Nova Scotia, Canada

Keywords: Multivariate statistical inference, elliptically contoured distributions, quadratic forms, multivariate normal distribution, parameter estimation, tests of hypotheses, linear models, multiple regression, partial correlation, multivariate analysis of variance, discriminant analysis, covariance models, principal components, factor analysis.

Contents

1. Introduction
2. Multivariate Distributions
2.1 The Multivariate Normal Distribution
2.2. Elliptically Contoured Distributions
2.3 The Wishart Distribution
3. Parameter Estimation for a Multivariate Normal Population
4. Tests of Hypotheses for Mean Vectors and Covariance Matrices
5. The General Linear Hypothesis Model
5.1 Multiple Regression and Correlation
5.2 Multivariate Analysis of Variance
6. Discriminant Analysis
7. Principal Components
8. Factor Analysis
Glossary
Bibliography
Biographical Sketches

Summary

Decisions in various spheres of activity including those of environmental or geopolitical nature are often made on the basis of statistical analyses involving several variables. Certain basic results on multivariate distributions as well as some of the main techniques utilized in multivariate statistical inference are presented in this chapter. Some numerical examples illustrate the theory.

1. Introduction

Multivariate methods analyze several variables simultaneously, unlike the more familiar univariate or bivariate methods which deal with only one or two variables. A variable can be independent (or explanatory) in which case it is a quantity used to explain or
predict the values of other variables which are called dependent (or response) variables. Multivariate statistical inference is often based on a data table (also called a data matrix) consisting of rows and columns. In this chapter, the rows contain the variables and the columns represent the objects or individuals being studied. Many of the results derived in multivariate analysis rely on matrix algebra, and it is assumed that reader has some knowledge of linear algebra in addition to being familiar with the basic concepts of mathematical statistics.

Section 2 introduces certain distributional properties of multivariate normal and elliptically contoured random vectors; some basic results are also presented in connection with quadratic forms and the Wishart distribution is defined. Estimators of the mean and covariance matrix of a multivariate normal distribution are given in Section 3, while useful tests of hypotheses are enumerated in Section 4. Section 5 covers multiple regression and correlation as well as one- and two-way multivariate analysis of variance. Discriminant analysis is introduced in Section 6. Certain multivariate covariance structures are discussed in Section 7 and Section 8 which, respectively, deal with principal components and factor analysis.

2. Multivariate Distributions

The multivariate normal distribution is the most widely used distribution in multivariate statistical inference. Some of its main properties are summarized in Section 2.1 where certain distributional results on quadratic forms in normal random vectors are also given. The normal distribution belongs to the more general class of elliptically contoured distributions which is described in Section 2.2. A brief definition of the Wishart distribution is given in Section 2.3.

2.1 The Multivariate Normal Distribution

A $p$-variate random vector $\mathbf{X}$ is said to have a real nonsingular normal distribution with mean $\mu = \mathbb{E}(\mathbf{X})$ and real positive definite covariance matrix $\Sigma = \mathbb{E}[(\mathbf{X} - \mathbb{E}(\mathbf{X})) (\mathbf{X} - \mathbb{E}(\mathbf{X}))^\prime]$, if its density function is given by

$$f(\mathbf{x}) = \frac{\exp\{-(\mathbf{x} - \mu)^\prime \Sigma^{-1} (\mathbf{x} - \mu)/2\}}{(2\pi)^{p/2} |\Sigma|^{1/2}}$$

(2.1.1)

The notation $\mathbf{X} \sim \mathcal{N}_p (\mu, \Sigma)$ will indicate that the $p$-variate random $\mathbf{X}$ is normally distributed with the parameters $\mu$ and $\Sigma$. When $\mu = \mathbf{0}$ (the null vector) and $\Sigma = \mathbf{I}$ (where $\mathbf{I}$ is a diagonal matrix whose diagonal elements are all equal to one), $\mathbf{X}$ is said to have a standard normal distribution. If the covariance matrix $\Sigma$ of a random vector $\mathbf{X}$ is diagonal, then the components of $\mathbf{X}$ are independently distributed.

Let $\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}$ where $\mathbf{X}_1$ has $q$ components and $\mathbf{X}_2$ has $p-q$ components, and $\mathbf{X} \sim \mathcal{N}_p (\mu, \Sigma)$ with
\[ \mathbf{\mu} = \begin{pmatrix} \mathbf{\mu}_1 \\ \mathbf{\mu}_2 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} ; \]

then \( X_1 \) and \( X_2 \) are independently and normally distributed with means \( \mu_1 \) and \( \mu_2 \) and covariance matrices \( \Sigma_{11} \) and \( \Sigma_{22} \), respectively, whenever \( \Sigma_{12} = 0 \) and \( \Sigma_{21} = 0 \). \( O \) denoting the null matrix whose elements are all equal to zero. Moreover, letting \( Y_1 = X_1 \) and \( Y_2 = X_2 - \Sigma_{21} \Sigma_{11}^{-1} X_1 \), it can be shown that \( Y_1 \) and \( Y_2 \) are independently distributed with \( Y_1 \sim N_q (\mu_1, \Sigma_{11}) \) and \( Y_2 \sim N_{p-q} (\mu_2 - \Sigma_{21} \Sigma_{11}^{-1} \mu_1, \Sigma_{22.1}) \) where \( \Sigma_{22.1} = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \). It follows that the conditional distribution of \( X_2 \) given \( X_1 = x \) is a multivariate normal with mean \( \mu_2 + \Sigma_{21} \Sigma_{11}^{-1} (x_1 - \mu_1) \) and covariance matrix \( \Sigma_{22.1} \).

Let \( A \) be a \( q \times p \) matrix of rank \( q \) with \( q \times p \), \( b \) be a \( q \)-dimensional vector of constants and \( X \sim N_p (\mu, \Sigma) \), then \( AX + b \sim N_q (A\mu + b, A\Sigma A') \). As an application, letting \( p = q \), \( A = \Sigma^{-1/2} \) and \( b = -\Sigma^{-1/2} \mu \) where \( \Sigma^{-1/2} \) denotes the symmetric square root of \( \Sigma^{-1} \), one can standardize the vector \( X \sim N_p (\mu, \Sigma) \) by means of the transformation \( Z = \Sigma^{-1/2} (X - \mu) \), and then \( Z \sim N_p (0, I) \). When \( q=1 \) (the case of a linear combination of the components of \( X \)), one has \( a'X \sim N(a'\mu, a'\Sigma a) \) for any real vector \( a \). Moreover, if \( b'X \) is univariate normal for every real vector \( b \), then \( X \) is distributed as a multivariate normal vector; this is a characterization of the multivariate normal distribution.

If \( \Sigma \), the covariance matrix of \( X \), has rank \( r \leq p \), then one can write \( \Sigma = BB' \), where \( B \) is \( p \times r \). Whether \( \Sigma \) is singular or nonsingular, there exists a standard normal vector \( Y \sim N_r (0, I) \) such that

\[ X = \mu + BY \quad (2.1.2) \]

In the singular normal case, the density of \( X \) does not exist but all the properties of \( X \) can be studied through the vector \( Y \) of (2.1.2). In the nonsingular case, the matrix \( B \) in (2.1.2) has dimension \( p \times p \) and is of rank \( p \).

The characteristic function of \( X \sim N_p (\mu, \Sigma) \) is

\[ M_X(t) = E(\exp \{it'X\}) = \exp \{it'\mu - (t'\Sigma t)/2\} \quad (2.1.3) \]

where \( t \) is a \( p \)-dimensional real vector and \( i = \sqrt{-1} \).

### Quadratic Forms in Normal Vectors

Consider the quadratic form \( X' BX \), where \( X \sim N_p (0, \Sigma) \) with \( \Sigma > 0 \) and \( B \) is a real symmetric matrix. Writing \( X \) as \( \Sigma^{1/2} Z \) where \( \Sigma^{1/2} \) is such that \( \Sigma^{1/2} \Sigma^{1/2} = \Sigma \) and \( Z \sim
\(\mathbf{X}'\mathbf{B}\mathbf{X} = \mathbf{Z}'(\Sigma^{1/2}\mathbf{B}\Sigma^{1/2})\mathbf{Z}\). Let \(\mathbf{Q}\) be an orthogonal matrix (with the property that \(\mathbf{Q}'\mathbf{Q} = \mathbf{Q}\mathbf{Q}' = \mathbf{I}\)) which diagonalizes \(\Sigma^{1/2}\mathbf{B}\Sigma^{1/2}\). Then \(\mathbf{Q}'(\Sigma^{1/2}\mathbf{B}\Sigma^{1/2})\mathbf{Q} = \mathbf{A}\), where \(\mathbf{A}\) is a diagonal matrix whose diagonal elements \(\lambda_i\) are the characteristic roots of \(\Sigma^{1/2}\mathbf{B}\Sigma^{1/2}\); and \(\mathbf{X}'\mathbf{B}\mathbf{X} = \mathbf{Z}'\mathbf{Q}\mathbf{A}\mathbf{Q}'\mathbf{Z}\). Note that on letting \(\mathbf{W} = \mathbf{Q}'\mathbf{Z}\), one has \(\mathbf{W} \sim N_p(0, \mathbf{Q}'\mathbf{I}\mathbf{Q}) \sim N_p(0, \mathbf{I})\). Thus,

\[
\mathbf{X}'\mathbf{B}\mathbf{X} = \mathbf{W}'\mathbf{A}\mathbf{W} = (W_1 \ W_2 \ldots \ W_p) \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ \vdots & \vdots & \ddots \\ 0 & 0 & \lambda_p \end{pmatrix} \begin{pmatrix} W_1 \\ W_2 \\ \vdots \\ W_p \end{pmatrix} = \sum_{i=1}^{p} \lambda_i W_i^2,
\]

where the \(W_i\)'s are independent standard normal variables, that is, the quadratic form \(\mathbf{X}'\mathbf{B}\mathbf{X}\) is distributed as a linear combination of independent chi-square variables each having one degree of freedom.

We now state two basic results in connection with quadratic forms.

1. Let \(\mathbf{X} \sim N_p(0, \mathbf{I})\); then \(\mathbf{X}'\mathbf{B}\mathbf{X} \sim \chi^2_r\) if and only if \(\mathbf{B} = \mathbf{B}^2\) (that is, \(\mathbf{B}\) is idempotent) and the rank of \(\mathbf{B}\) is \(r\).

2. Let \(\mathbf{Q}_1 = \mathbf{X}'\mathbf{A}\mathbf{X}\) and \(\mathbf{Q}_2 = \mathbf{X}'\mathbf{B}\mathbf{X}\) where \(\mathbf{X} \sim N_p(0, \mathbf{I})\), \(\mathbf{A} = \mathbf{A}'\) and \(\mathbf{B} = \mathbf{B}'\); then \(\mathbf{Q}_1\) and \(\mathbf{Q}_2\) are independently distributed if and only if \(\mathbf{AB} = \mathbf{BA} = \mathbf{O}\) (Craig’s theorem).

### 2.2. Elliptically Contoured Distributions

We shall first present some of the main properties characterizing the class of elliptically contoured distributions via the multivariate normal distribution.

Consider the equation \(f(\mathbf{x}) = c_1\) where \(c_1\) is a constant and \(f(\mathbf{x})\) is the density function of a multivariate normal distribution given in Equation 2.1.1. Then

\[
-\log f(\mathbf{x}) = -\log c_1 = (\mathbf{x} - \mathbf{\mu})'\Sigma^{-1}(\mathbf{x} - \mathbf{\mu}) = c
\]

where \(c\) is a constant. Thus the contours of constant density are ellipsoids whenever \(\mathbf{X} \sim N_p(\mathbf{\mu}, \Sigma)\) when \(\mathbf{\mu} = \mathbf{0}\) and \(\Sigma = \mathbf{I}\), the identity matrix, these ellipsoids are spheres centered at the origin and \(\mathbf{X}\) is said to have a spherical distribution. Moreover, the characteristic function of \(\mathbf{Y} \sim N_p(0, \mathbf{I})\) as well that of \(\mathbf{Z} = \mathbf{PY}\), where \(\mathbf{P}\) is an orthogonal matrix, are one and the same, which means that there is invariance under orthogonal transformations. If the factor \(\exp\{-\frac{1}{2}(t'\Sigma t)\}\) of (2.1.3) is replaced by a general nonnegative function of \(t'\Sigma t\), the resulting distribution is said to belong to the class of elliptically contoured distributions. Elliptically contoured and spherical distributions are formally defined below.
Let $\mu$ be a $p$-dimensional real vector, $\Sigma$ be a $p \times p$ nonnegative definite matrix and $\xi(.)$ be a nonnegative function; then the $p$-dimensional vector $X$ is said to have an elliptical or elliptically contoured distribution if its characteristic function $\phi(t)$ can be expressed as $\exp(it'\mu)\xi(t'\Sigma t)$, and we write $X \sim C_p (\xi; \mu, \Sigma)$. If some components of $t$ in the characteristic function of $X$ are set to zero, the resulting characteristic function still has the same form, which means that if $X$ has an elliptically contoured distribution then all the marginal distributions are also elliptical. When $\mu$ is the null vector and $\Sigma$ is the identity matrix of order $p$, the notation $X \sim C_p (\xi; \mu, \Sigma)$ is shortened to $X \sim S_p (\xi)$, and $X$ is said to have a spherical or spherically symmetric distribution.

Let $Y$ be spherically distributed with p.d.f. $g(y'y)$ and $X$ be distributed as $\mu + BY$ where $\mu$ is a $p$-dimensional vector and $B$ is a $p \times p$ matrix such that $\Sigma = BB'$; then $X$ has the p.d.f. $|\Sigma|^{-1/2} g((x-\mu)'\Sigma^{-1}(x-\mu))$. Clearly, the contours of constant density of $X$ are ellipsoids as in the multivariate normal case.

Elliptically contoured distributions can also be defined as follows: Let $\Sigma$ be $p \times p$ matrix of rank $q \leq p$; then there exists a $q \times p$ matrix $L$ such that $\Sigma = LL'$. Now, letting $W$ be a nonnegative random variable which is identically distributed in every direction along radii from the point $\mu$ and $U(\alpha)$ denote a random vector which is uniformly distributed on the unit sphere in $\mathbb{R}^q$ and whose distribution is independent of that of $W$, one has $X \sim C_p (\xi; \mu, \Sigma)$ if and only if $X \sim \mu + L' W U(\alpha)$ and $F(w)$, the distribution function of $W$, is such that

$$
\xi(t't) = E_w [E_{w'|w=w}(\exp\{it'wU(\alpha)\})] = \int E(\exp\{iwt'U(\alpha)\})dF(w)
$$

where $\xi(t't)$ is the characteristic function of $WU(\alpha) \sim S_q (\xi)$.

2.3 The Wishart Distribution

Letting $X_1, X_2, \ldots, X_n$ be a random sample of size $n$ from a $N_p (\mu, \Sigma)$ population,

$$
X = \begin{pmatrix}
X_1 \\
X_2 \\
\vdots \\
X_n
\end{pmatrix}
$$

and

$$
A = \begin{pmatrix}
a_{11} & a_{12} & \cdots & a_{1p} \\
a_{21} & a_{22} & \cdots & a_{2p} \\
\vdots & \vdots & \ddots & \vdots \\
a_{p1} & a_{p2} & \cdots & a_{pp}
\end{pmatrix}
$$

the random matrix $A$ is said to follow a Wishart distribution with $n$ degrees of freedom and covariance parameter $\Sigma$, and we write $A \sim W_p (n, \Sigma)$.

3. Parameter Estimation for a Multivariate Normal Population
A multivariate normal vector $X \sim N_p (\mu, \Sigma)$ is specified completely by its mean $\mu$ and covariance matrix $\Sigma$. Given a random sample $X_1, X_2, \ldots, X_N$ with $N > p$ from this distribution, the maximum likelihood estimators of $\mu$ and $\Sigma$ are obtained by maximizing the likelihood function of the sample, which is
\[
\exp\left\{ -\frac{1}{2} \text{tr}(\Sigma^{-1} [\sum_{i=1}^{N} (x_i - \mu)(x_i - \mu)']) \right\}
\]
\[
\frac{1}{(2\pi)^{pN/2} |\Sigma|^{N/2}}.
\]
They are respectively
\[
\hat{\mu} = \bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i \quad \text{and} \quad \hat{\Sigma} = \frac{1}{N} \sum_{i=1}^{N} (X_i - \bar{X})(X_i - \bar{X})'.
\]
The sample mean $\bar{X}$ is normally distributed as $N_p (\mu, \Sigma/N)$, and so $N(\bar{X} - \mu)' \Sigma^{-1} (\bar{X} - \mu)$ has a chi-square distribution with $p$ degrees of freedom. It can also be shown that $N\hat{\Sigma}$ is distributed as $\sum_{i=1}^{N-1} Z_i Z_i'$ where the $Z_i$'s are independently distributed as $N_p(0, \Sigma)$, $i = 1, 2, \ldots, N-1$. Consequently, letting the sample covariance $S = \sum_{i=1}^{N} (X_i - \bar{X})(X_i - \bar{X})'(N-1)$, $(N-1)S = N\hat{\Sigma}$ is seen to be distributed as a Wishart random matrix with $(N-1)$ degrees of freedom. The estimators $\bar{X}$ and $S$ are independently distributed and respectively unbiased for $\mu$ and $\Sigma$.

4. Tests of Hypotheses for Mean Vectors and Covariance Matrices

Some useful test statistics for the mean vectors and covariance matrices of multivariate normal populations are enumerated below.

1. Testing that a mean is equal to a given vector when the covariance matrix is known.
Let $X_1, \ldots, X_N$ be a random sample from $N_p (\mu, \Sigma)$ and $\Sigma$ be a known covariance matrix; then a rejection region of size $\alpha$ to test the hypothesis $\mu = \mu_0$, where $\mu_0$ is a specified vector, is given by
\[
N(\bar{X} - \mu_0)' \Sigma^{-1} (\bar{X} - \mu_0) > \chi_p^2 (\alpha)
\]
where $\chi_p^2 (\alpha)$ is a critical value such that $P[\chi_p^2 > \chi_p^2 (\alpha)] = \alpha$, and a 100 $(1 - \alpha)%$ confidence region for the mean vector $\mu$ is specified by the set of points $m$ satisfying the following inequality:
\[
N(\bar{X} - m)' \Sigma^{-1} (\bar{X} - m) \leq \chi_p^2 (\alpha).
\]

2. Testing equality of the mean vectors in two populations with known common covariance matrix.
Let $X_1, \ldots, X_N$ and $Y_1, \ldots, Y_M$ be random samples from $N_p(\mu_1, \Sigma)$ and $N_p(\mu_2, \Sigma)$, respectively, with sample means $\overline{X}$ and $\overline{Y}$, and assume that $\Sigma$ is known. Then, $(\overline{X} - \overline{Y})$ is distributed as $N_p(\mu_1 - \mu_2, [(1/N) + (1/M)]\Sigma)$ and a rejection region for testing the hypothesis $\mu_1 = \mu_2$, is given by

$$\frac{NM}{N + M} (\overline{X} - \overline{Y})' \Sigma^{-1} (\overline{X} - \overline{Y}) > \chi_p^2(\alpha).$$

3. Testing that a mean is equal to a given vector when the covariance matrix is unknown

The likelihood ratio test of the hypothesis $\mu = \mu_0$ for a $N_p(\mu, \Sigma)$ population is based on $T^2 = N(\overline{X} - \mu_0)' S^{-1} (\overline{X} - \mu_0)$, where $\overline{X}$ and $S$ are respectively the sample mean and the sample covariance. A 100 (1 - $\alpha$)% confidence region for the mean vector $\mu$ is given by

$$N(\overline{X} - m)' S^{-1} (\overline{X} - m) \leq \frac{(N-1)p}{(N-p)} F_{p,N-p}(\alpha)$$

as an inequality with respect to $m$ where $F_{p,N-p}(\alpha)$ is such that $P(F_{p,N-p} > F_{p,N-p}(\alpha)) = \alpha$, $F_{p,N-p}$ denoting the $F$ distribution with $p$ and $N-p$ degrees of freedom.

4. Testing equality of means in two populations with unknown common covariance matrix

Let $X_1, \ldots, X_N$ and $Y_1, \ldots, Y_M$ be random samples from $N_p(\mu_1, \Sigma)$ and $N_p(\mu_2, \Sigma)$, respectively, $\Sigma$ being unknown, $T^2 = NM (\overline{X} - \overline{Y})' S^{-1} (\overline{X} - \overline{Y})/(N+M)$, and $S = \frac{1}{N + M - 2} \left\{ \sum_{i=1}^{N} (X_i - \overline{X})(X_i - \overline{X})' + \sum_{i=1}^{M} (Y_i - \overline{Y})(Y_i - \overline{Y})' \right\}$. A rejection region for testing the hypothesis $\mu_1 = \mu_2$, is given by

$$T^2 > \frac{(N + M - 2)p}{N + M} F_{p,N+M-p-1}(\alpha).$$

5. Testing the hypothesis that a mean vector and a covariance matrix are equal to a given vector and matrix.

Let $X_1, X_2, \ldots, X_N$ be a random sample from $N_p(\mu, \Sigma)$ where $\Sigma$ is positive definite covariance matrix. The likelihood ratio criterion for testing the hypothesis $H_0$: $\mu = \mu_0$, $\Sigma = \Sigma_0$, where $\mu_0$ and $\Sigma_0$ are given, is
\[
\lambda = \left( \frac{e}{N} \right)^{\frac{1}{2}} \left| A \Sigma_0^{-1} \right|^{\frac{1}{2}} N \exp \left\{ - \frac{1}{2} \left[ \text{tr}(A \Sigma_0^{-1}) + N(\bar{X} - \mu_0)' \Sigma_0^{-1}(\bar{X} - \mu_0) \right] \right\}
\]

where \( A = \sum_{i=1}^{N} (X_i - \bar{X})(X_i - \bar{X})' \). Under the null hypothesis, \(-2\log \lambda\) is asymptotically distributed as a chi-square random variable with \( p(p+1)/2 + p \) degrees of freedom.

6. Testing the hypothesis that a covariance matrix is equal to a given matrix.

Let \( X_1, X_2, \ldots, X_N \) be a random sample from \( N_p(\mu, \Sigma) \) where \( \Sigma \) is a positive definite covariance matrix. The likelihood ratio criterion for testing the hypothesis \( H_0: \Sigma = \Sigma_0 \), where \( \Sigma_0 \) is a given covariance matrix, is

\[
\lambda = \left( \frac{e}{N} \right)^{\frac{1}{2}} \left| A \Sigma_0^{-1} \right|^{\frac{1}{2}} N \exp \left\{ - \frac{1}{2} \left[ \text{tr}(A \Sigma_0^{-1}) \right] \right\}
\]

where \( A = \sum_{i=1}^{N} (X_i - \bar{X})(X_i - \bar{X})' \). Under the null hypothesis, \(-2\log \lambda\) is asymptotically distributed as a chi-square random variable with \( p(p+1)/2 + p \) degrees of freedom.

7. Testing the hypothesis that a covariance matrix is proportional to an identity matrix.

Let \( X_1, X_2, \ldots, X_N \) be a random sample from \( N_p(\mu, \Sigma) \) where \( \Sigma \) is a positive definite covariance matrix. The likelihood ratio criterion for testing the hypothesis \( H_0: \Sigma = \sigma^2 I \), where \( \sigma^2 \) is not specified, is

\[
\lambda = \left( \frac{e}{N} \right)^{\frac{1}{2}} \left| A \right|^{\frac{1}{2}} N \exp \left\{ - \frac{1}{2} \left[ \text{tr}(A) \right] \right\}
\]

\[
(\text{tr}(A)/p)^{\frac{1}{2}} N \exp \left\{ - \frac{1}{2} \left[ \text{tr}(A) \right] \right\}
\]

where \( A = \sum_{i=1}^{N} (X_i - \bar{X})(X_i - \bar{X})' \). Under the null hypothesis, \(-2\log \lambda\) is asymptotically distributed as a chi-square random variable with \( p(p+1)/2 - p \) degrees of freedom. This test is called the sphericity test. It determines whether the components of a random vector are uncorrelated and they all have equal variances.
Bibliography


Biographical Sketches
Serge B. Provost is Professor in the Department of Statistical and Actuarial Sciences at the University of Western Ontario. He received his B.Sc and M.Sc degrees from l’université de Montréal and his Ph.D. degree in Mathematics and Statistics from McGill University in 1984. He is a Fellow of the Royal Statistical Society, a Chartered Statistician, an elected member of the International Statistical Institute, and an Associate of the Society of Actuaries. He is a member of the American Statistical Association, the Institute of Mathematical Statistics and the Statistical Society of Canada. He also belongs to the Education sections of the American Statistical Association, the International Statistical Institute and the Society of Actuaries. He has supervised seventeen Master's students and three Ph.D. students. He edited issues of the International Journal of Mathematical and Statistical Sciences and the Pakistan Journal of Statistics, co-edited three books and is Editor-at-Large for the Statistics series published by Marcel Dekker, Inc. He was an invited research scholar at the Centre for Mathematical Studies, Trivandrum, India, in December 2000 and the Korean Advanced Institute for Science and Technology, Taejon, Korea, in August 2001. He also served on the Bilingualism Committee of the Statistical Society of Canada and is currently an academic consultant for STATLAB, the department statistical consulting unit. He is the author of two research monographs, four encyclopedia articles and some thirty-five papers published in peer-reviewed publications. The joint paper with E. M. Rudiuk entitled "The sampling distribution of the serial correlation coefficient" and published in 1995 in The American Journal of Mathematical and Management Sciences was awarded the Jacob Wolfowitz Prize in 1997. His research interests include distribution theory, time series, density estimation, order statistics and multivariate analysis with applications to the environmental sciences, econometrics and engineering, among other disciplines.

Edmund M. Rudiuk received an M.Sc. degree in Mathematics from the University of Maria Curie-Skłodowska, Lublin, Poland as well as M.Sc. and Ph.D. degrees in Statistics from the University of Western Ontario in 1992 and 1993, respectively. He is currently an Associate Professor of Statistics at the University College of Cape Breton, Sydney, Canada. His research interests include the distribution of ratios of quadratic forms and in particular that of the sample serial correlation coefficient as well as applications in distribution theory of the inverse Mellin transform technique and the generalized hypergeometric functions. On the practical side, he was involved in the analysis of cancer data and of pollution data at the tar ponds and coke oven site in Sydney, Cape Breton. In 1997, he was awarded the Jacob Wolfowitz Prize by the Editorial Board of the American Journal of Mathematical and Management Sciences.