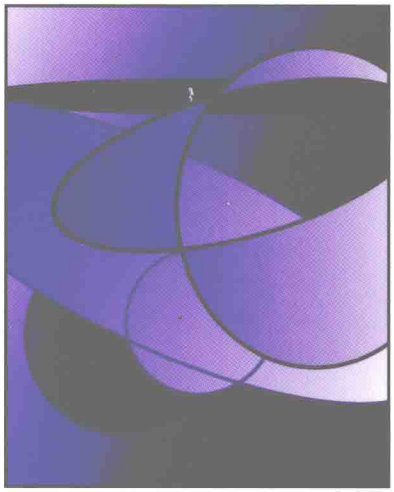

Multivariate Statistical Methods

FOURTH EDITION

Donald F. Morrison

Multivariate Statistical Methods

Fourth Edition

Donald F. Morrison
The Wharton School
University of Pennsylvania

THOMSON
BROOKS/COLE
Australia • Canada • Mexico • Singapore • Spain • United Kingdom • United States

THOMSON
BROOKS/COLE

Acquiring Editor: *Carolyn C. Crockett*
Assistant Editor: *Ann Day*
Editorial Assistant: *Rhonda Letts*
Technology Project Manager: *Burke Taft*
Marketing Manager: *Tom Ziolkowski*
Advertising Project Manager: *Nathaniel Bergson-Michelson*
Project Manager, Editorial Production: *Kelsey McGee*

Manufacturing Buyer: *Emma Claydon*
Permissions Editor: *Stephanie Lee*
Typesetter: *International Typesetting and Composition*
Cover Designer: *Jennifer Macres*
Cover Printing, Printing and Binding: *Quebecor World/Kingsport*

Printed in the United States of America
1 2 3 4 5 6 7 08 07 06 05 04

For more information about our products, contact us at:
Thomson Learning Academic Resource Center
1-800-423-0563

For permission to use material from this text, contact us by:

Web: http://www.thomsonrights.com

ISBN 0-534-38778-0

Brooks/Cole Thomson Learning
10 Davis Drive
Belmont, CA 94002
USA

Asia
Thomson Learning
5 Shenton Way #01-01
UIC Building
Singapore 068808

Australia/New Zealand
Thomson Learning
102 Dodds Street
Southbank, Victoria 3006
Australia

Canada
Nelson
1120 Birchmount Road
Toronto, Ontario M1K 5G4
Canada

Europe/Middle East/Africa
Thomson Learning
High Holborn House
50/51 Bedford Row
London WC1R 4LR
United Kingdom

For Phyllis

Preface

This is the fourth edition of *Multivariate Statistical Methods*. In writing the first edition, my original motivation came from the kinds of statistical problems brought by investigators in the life, medical, and behavioral sciences at the National Institute of Mental Health, and the need for a text and reference source which did not presuppose several courses in statistical theory. Subsequently the book was used as the basis for a one-semester course in multivariate methods for thirty-plus years at the Wharton School of the University of Pennsylvania.

In planning the revised volume I concluded that the material of the previous Chapter 1, an overview of traditional univariate statistics, and Chapter 2, a summary of useful matrix algebra concepts, were readily available in many other texts, and should be omitted. Hence, those chapters of the third edition have been scanned to a website for the current edition that is maintained by Duxbury Press.

Certain themes continue in this edition: The extension of univariate tests on the mean to multidimensional mean vectors through the Hotelling T^2 statistic and the multivariate analysis of variance (MANOVA), classification by various discrimination measures, and the dissection of covariance structure by principal component and factor analysis. Wherever possible, the tests are followed by simultaneous inferential procedures. For that reason I still prefer S. N. Roy's union-intersection development of the tests, and its natural concomitant of simultaneous confidence intervals for any requisite multiple comparisons.

I have attempted to reference many of the relevant contributions to multivariate analysis from the mainstream journals over the past fifteen years in this edition, especially in such areas as the analysis of repeated measures data. Additional exercises using actual data sets have been included, particularly in Chapters 2, 3, and 6. A variety of disciplines is represented by those data. A number of examples arose from my participation in the Cerebrovascular Research Center at the Medical School of the University of Pennsylvania, and I am most grateful to Dr. Martin Reivich and his colleagues for that stimulating association and the use of their data here. Other sets, *e.g.*, course and instructor ratings, may be more mundane, but still nicely illustrate certain multivariate concepts and methods.

As in the earlier editions no single statistical software system is used for implementing the multivariate analysis. I have found MINITAB convenient and user-friendly for MANOVA, discrimination and classification, extracting principal components, and factor analysis, and I am grateful to MINITAB, Inc., for providing copies of its software under its Author Assistance Program. In other examples I have used APL for evaluating linear and quadratic functions of data, and especially its EIG and SYMEIG functions for computing the characteristic roots and vectors of matrices.

I am greatly indebted to Carolyn Crockett, Senior Acquisitions Editor of Duxbury Press, for her support of this project. Kelsey McGee at Duxbury was very helpful for her encouragement and

interest in the book during its preparation. Jennifer Jenkins and Rhonda Letts at Duxbury were continually supportive during the writing and composition phases.

The initial composition of this book in LaTex pages would not been possible without an intervention of my wife, Phyllis Morrison. Before a talk by Professor Richard Kadison of the Penn Mathematics Department to the University Women's club, she mentioned my project. Professor Kadison introduced me to his graduate student, Junhao Shen, who agreed to undertake the lengthy task of composing the book by the TeX language. I am greatly indebted to Mr. Shen for his steadfast work in producing the pages and their numerous tables and mathematical displays. Any errors that might appear are, of course, my own responsibility.

Donald F. Morrison

Contents

Chapter 1

Samples from the Multivariate Normal Population

1.1 Introduction

In this chapter we shall extend the concept of a continuous random variable to variates defined in several dimensions. We shall concentrate our attention on the multivariate generalization of normally distributed random variables. The parameters of that multinormal distribution will be related to multiple and partial correlation measures for describing relations among the dimensions. We shall consider means of estimating the multinormal parameters from random samples of observations on the variates, as well as the sampling distributions of estimates and related statistics. In the methods for hypothesis tests and confidence statements on parametric functions we shall emphasize procedures which control error rates for several simultaneous inferences. Some attention will be given to estimation when observations are missing at random in the data.

1.2 Why Do We Need Multivariate Methods?

Univariate statistical methods deal with single variables: Quantitative Graduate Record Examination scores for students matriculating in a university program, blood glucose levels in the control group of a nutrition study at a particular time, or the total score of a cognitive ability test given to twenty-year-old women as part of a longitudinal investigation of human aging. In each case we wish to make statistical inferences about the population distribution of the variable, set confidence intervals for its parameters, and perhaps test hypotheses about the values of the parameters. Our view of each variable is one-dimensional.

Multivariate analysis considers several variables at once. For example, we might wish to investigate the variation in the rate at which the human brain metabolizes glucose at different anatomical regions in a sample of young normal male subjects. The cerebral glucose metabolic rate, denoted by CMRgl, can be measured by positron emission tomography (PET) scans of the brain when the subject has received an injection of a radioactive tracer substance, such as Fluorodeoxyglucose-18. Then, for p anatomical regions of interest or coordinate locations in the brain the measured CMRgl values can be represented by the row vector of observations

$$\mathbf{X}' = [x_1, \ldots, x_p]$$

If we have a sample of N subjects the data can be represented conveniently in matrix form as

$$\mathbf{X} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ & \cdots & \\ x_{N1} & \cdots & x_{Np} \end{bmatrix}$$

Each row contains the observations from a subject, while the columns give CMRgl for a particular location in the brain. In the algebraic sense, we have extended one-dimensional data to vectors with p components.

If we have a random sample of subjects, and if the assumption of a multidimensional normal distribution holds, we can answer these questions about the parameters of the distribution:

1. Are the CMRgl population means simultaneously equal to p specified values?
2. Is the mean CMRgl the same at the p locations in the brain?
3. What are the confidence intervals for the p population CMRgl means?
4. What can we infer about the variability and dependence structures of the p variables? Does the structure have a pattern that might imply simpler hidden structures?

Multivariate analysis essentially extends univariate tests and confidence intervals for parameters to vectors or matrices of parameters.

Frequently multivariate data are collected in more complex designs than a simple random sample, and the univariate methods based on those experimental designs can be extended to observation vectors to produce tests for the equality of two or more mean vectors. For example, in the case of CMRgl measured by PET scans we might have two separate samples of young (under age 40) and old (age 40 and over) subjects. The regions of interest (ROIs) in the brain are also designated by left or right hemisphere. If each anatomic structure has corresponding left and right locations, the metabolic rates might be arranged as in Table 1.1, where the observations are denoted by a generic x to avoid the unnecessary complications of subscripts.

Multivariate statistical theory provides methods for making these inferences about the distribution of old and young CMRgl variables:

1. Are the CMRgl means different for the old and young populations?
2. Do the left and right hemispheres differ in the CMRgl means?

TABLE 1.1
Cerebral Metabolic Rates Cross-Classified by Age and Hemisphere

		Hemisphere	
		Left Region	Right Region
Age	**Subject**	$1 \ldots p$	$1 \ldots p$
Old	1	$x \ldots x$	$x \ldots x$
	N_1	$x \ldots x$	$x \ldots x$
Young	1	$x \ldots x$	$x \ldots x$
	N_2	$x \ldots x$	$x \ldots x$

3. Are the left-right hemisphere differences the same for the old and young populations?
4. Do the regions have the same CMRgl means, either aggregated over age and hemisphere, or separately within those classifications?
5. Are mean linear functions (*e.g.*, the average) the same for the old and young populations, or for the left and right hemispheres?
6. What are the properties of the variability and dependence structures for the age and hemisphere subgroups?

For a single CMRgl observation we could answer those questions by t and F tests, and the analysis of variance. For multivariate data those methods have been generalized to hypotheses tests on vectors of means, or multivariate analysis of variance for vector-valued observations. In the forthcoming sections of this book we shall describe those statistical methods. In the next section we shall begin by developing some of their underlying mathematical assumptions.

1.3 Multidimensional Random Variables

Let us define the p-dimensional random variable $\mathbf{X}$ as the vector

$$\mathbf{X}' = [X_1, \ldots, X_p] \tag{1}$$

whose elements are continuous unidimensional random variables with density functions $f_1(x_1), \ldots, f_p(x_p)$ and distribution functions $F_1(x_1), \ldots, F_p(x_p)$. In like manner $\mathbf{X}$ has the *joint distribution function*

$$F_1(x_1, \ldots, x_p) = P(X_1 \leq x_1, \ldots, X_p \leq x_p) \tag{2}$$

If that function is *absolutely continuous*, it is possible to write

$$F(x_1, \ldots, x_p) = \int_{-\infty}^{x_p} \cdots \int_{-\infty}^{x_1} f(u_1, \ldots, u_p)\, du_1 \ldots du_p \tag{3}$$

where $f(x_1, \ldots, x_p)$ is the *joint density function* of the elements of $\mathbf{X}$. If those quantities are *independent* random variables,

$$\begin{aligned} f(x_1, \ldots, x_p) &= f_1(x_1) \ldots f_p(x_p) \\ F(x_1, \ldots, x_p) &= F_1(x_1) \ldots F_p(x_p) \end{aligned} \tag{4}$$

Conversely, such factorizations of the density and distribution functions imply that the X_i are independent variates. Through the assumption of independence the forms of the joint distribution and density have been exceedingly simplified, and it is for this reason that classical statistical methods require random samples of observations to permit the various joint distributions to have this product form. However, the multivariate statistical models we shall encounter in this text will nearly always assume that the elements of the random vector are dependent, and the resulting analytical procedures and distribution theory will be constructed to be valid in the presence of this dependence. Still, as in the univariate case, we shall require that successive sample observation vectors from the multidimensional population have been drawn in such a way that they can be construed as realizations of independent random vectors.

The joint density of any subset of the elements of $\mathbf{X}$ is found by integrating the original joint density over the domain of the variates not in the subset. If the variates have been numbered conveniently, so that the subset consists of the elements $X_1, \ldots, X_p$, and the complement of the set contains $X_{p+1}, \ldots, X_{p+q}$, the joint density of the first variates is

$$g(x_1, \ldots, x_p) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_{p+q})\, dx_{p+1} \ldots dx_{p+q} \tag{5}$$

and the joint distribution function can be computed by setting the variates of the second subset equal to their upper limits in the original joint distribution function:

$$\begin{aligned} G(x_1, \ldots, x_p) &= P(X_1 \leq x_1, \ldots, X_p \leq x_p) \\ &= F(x_1, \ldots, x_p, \infty, \ldots, \infty) \end{aligned} \tag{6}$$

We note in particular that the marginal density of any single element of $\mathbf{X}$ is

$$f_i(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_p)\, dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_p \tag{7}$$

In such multiple integrals as (5) and (7) it is essential to recognize that the density functions have always been properly defined so that the limits of integration can be stated formally as $-\infty$ and ∞. For example, if Y_1 and Y_2 are independent variates with the common density and distribution function $g(y)$ and $G(y)$, respectively, the joint density of the new random variables

$$\begin{aligned} X_1 &= \text{smaller of } (Y_1, Y_2) \\ X_2 &= \text{larger of } (Y_1, Y_2) \end{aligned} \tag{8}$$

can be shown to be

$$f(x_1, x_2) = \begin{cases} 2g(x_1)g(x_2) & -\infty < x_1 \leq x_2 < \infty \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

The marginal densities are

$$\begin{aligned} f_1(x_1) &= \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_2 \\ &= \int_{-\infty}^{x_1} 0 dx_2 + 2\int_{x_1}^{\infty} g(x_1)g(x_2)\, dx_2 \\ &= 2g(x_1)[1 - G(x_1)] \qquad -\infty < x_1 < \infty \end{aligned} \tag{10}$$

$$\begin{aligned} f_2(x_2) &= \int_{-\infty}^{\infty} f(x_1, x_2) dx_1 \\ &= 2\int_{-\infty}^{x_2} g(x_1)g(x_2) dx_1 + \int_{x_2}^{\infty} 0\, dx_1 \\ &= 2g(x_2)G(x_2) \qquad -\infty < x_2 < \infty \end{aligned} \tag{11}$$

We observe in passing that while $f(x_1, x_2)$ factored into the product of two densitylike functions, the variates X_1 and X_2 are *not* independent, for these factors are not the marginal densities (10) and (11).

Conditional Distributions

It is frequently necessary in multivariate analysis to know the distribution of one set of random variables, given that the variates of a second group have been set equal to specified constant values or have been constrained to lie in some subregion of their complete space. Such distributions and density functions are called *conditional*. The density function of the conditional distribution of $X_1, \ldots, X_p$ given $X_{p+1} = x_{p+1}, \ldots, X_{p+q} = x_{p+q}$ can be shown to be

$$h(x_1, \ldots, x_p | x_{p+1}, \ldots, x_{p+q}) = f(x_1, \ldots, x_{p+q})/g(x_{p+1}, \ldots, x_{p+q}) \tag{12}$$

where $f(x_1, \ldots, x_{p+q})$ is the joint density of the complete set of $p + q$ variates and $g(x_{p+1}, \ldots, x_{p+q})$ is the positive joint density of the q fixed variates. If the two sets of variates are independent, factorization of the joint density implies that the conditional density of the first set is merely the joint density of those random variables. For any admissible set of values of the fixed variates the function (12) has all the properties of a density function, and if the original distribution function is absolutely continuous, the conditional distribution function can be computed in the usual manner as

$$F(x_1, \ldots, x_p | x_{p+1}, \ldots, x_{p+q}) = \frac{\int_{-\infty}^{x_p} \cdots \int_{-\infty}^{x_1} f(u_1, \ldots, u_p, x_{p+1}, \ldots, x_{p+q}) du_1 \ldots du_p}{g(x_{p+1}, \ldots, x_{p+q})} \tag{13}$$

The fact that $F(\infty, \ldots, \infty | x_{p+1}, \ldots, x_{p+q}) = 1$ should be immediately apparent from (5) and (13).

For an example let us refer to the random variables X_1 and X_2 whose joint density is specified by expression (9). The conditional density of X_1, given that X_2 has the value x_2, is

$$f(x_1|x_2) = \begin{cases} g(x_1)/G(x_2) & -\infty < x_1 \leq x_2 < \infty \\ 0 & \text{elsewhere} \end{cases} \tag{14}$$

The conditional distribution function is

$$F(x_1|x_2) = \begin{cases} G(x_1)/G(x_2) & -\infty < x_1 \leq x_2 < \infty \\ 1 & x_2 \leq x_1 < \infty \end{cases} \tag{15}$$

For all x_2, the conditional probability that X_1 will be at most equal to x_1 will be larger than the unconditional probability $G(x_1)$ of that event. Through the conditional distribution function we may use knowledge of the magnitude of X_2 to improve our prediction of the value of the first random variable.

Moments of Multidimensional Variates

The expected value of the random vector $\mathbf{X}$ is merely the vector of the expectations of its elements:

$$E(\mathbf{X}') = [E(X_1), \ldots, E(X_p)] \tag{16}$$

Similarly the expectation of a random matrix is the matrix of expected values of the random elements. For the generalization of the variance to multidimensional variates let us first define the covariance of the elements X_i and X_j of $\mathbf{X}$ as the product moment of those variates about their respective means:

$$\begin{aligned} \text{cov}\,(X_i, X_j) &= E\{[X_i - E(X_i)][X_j - E(X_j)]\} \\ &= E(X_i X_j) - [E(X_i)][E(X_j)] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_i x_j f_{ij}(x_i, x_j)dx_i dx_j - [E(X_i)][E(X_j)] \\ &= \sigma_{ij} \end{aligned} \tag{17}$$

where $f_{ij}(x_i, x_j)$ is the joint density of X_i and X_j. If $i = j$, the covariance is the variance of X_i, and we shall customarily write $\sigma_{ii} = \sigma_i^2$. The extension of the variance notion to the p-component random vector $\mathbf{X}$ is the matrix of variances and covariances

$$\begin{aligned} \text{cov}\,(\mathbf{X}, \mathbf{X}') &= E\{[\mathbf{X} - E(\mathbf{X})][\mathbf{X} - E(\mathbf{X})]'\} \\ &= \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \cdots & \cdots & \cdots \\ \sigma_{1p} & \cdots & \sigma_{pp} \end{bmatrix} \\ &= \boldsymbol{\Sigma} \end{aligned} \tag{18}$$

We shall call this symmetric matrix the *covariance matrix* of $\mathbf{X}$.

It is easily verified from the definition (17) that the covariance of two random variables is unchanged by shifts in the origins of those variates. Thus,

$$\text{cov}\,(X_i + a, X_j + b) = \text{cov}\,(X_i, X_j) \tag{19}$$

for all real constants a and b. Similarly, changes in scale of variates affect the covariance by the same factors:

$$\text{cov}\,(cX_i, dX_j) = cd\ \text{cov}\,(X_i, X_j) \tag{20}$$

This result leads to the expression for the variance of a linear compound $\mathbf{a}'\mathbf{X} = a_1X_1 + \cdots + a_pX_p$ of random variables:

$$\text{var}\,(\mathbf{a}'\mathbf{X}) = \sum_{i=1}^{p}\sum_{j=1}^{p} a_i a_j \sigma_{ij} = \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} \tag{21}$$

The covariance of the two linear compounds $\mathbf{a}'\mathbf{X}$ and $\mathbf{b}'\mathbf{X}$ in the same random variables is the bilinear form

$$\text{cov}\,(\mathbf{a}'\mathbf{X}, \mathbf{b}'\mathbf{X}) = \sum_{i=1}^{p}\sum_{j=1}^{p} a_i b_j \sigma_{ij} = \mathbf{a}'\boldsymbol{\Sigma}\mathbf{b} \tag{22}$$

More generally, if $\mathbf{A}$ and $\mathbf{B}$ are of dimensions $r \times p$ and $s \times p$, respectively, and contain real elements, the covariances of the transformed variates

$$\mathbf{Y} = \mathbf{A}\mathbf{X} \qquad \mathbf{Z} = \mathbf{B}\mathbf{X}$$

will be given by the matrices

$$\begin{aligned} \text{cov}\,(\mathbf{Y}, \mathbf{Y}') &= \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}' \\ \text{cov}\,(\mathbf{Z}, \mathbf{Z}') &= \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}' \\ \text{cov}\,(\mathbf{Y}, \mathbf{Z}') &= \mathbf{A}\boldsymbol{\Sigma}\mathbf{B}' \end{aligned} \tag{23}$$

The correlation coefficient of the variates X_i and X_j is defined as

$$\rho_{ij} = \frac{\text{cov}(X_i, X_j)}{\sqrt{\text{var}(X_i)\text{var}(X_j)}} \tag{24}$$

By the properties (19) and (20) it follows that the correlation is a pure-number invariant under changes of scale and origin of its variates. From the properties of the integrals defining the variance and covariance it can be shown that ρ cannot be less than -1 or greater than 1. If X_i and X_j are independently distributed, their covariance, and hence their correlation, is zero. However, the converse is not generally true, for it is possible to construct examples of highly dependent random variables whose correlation is zero.

Later we shall need the matrix of population correlations

$$\mathbf{P} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{12} & 1 & \cdots & \rho_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ \rho_{1p} & \rho_{2p} & \cdots & 1 \end{bmatrix} \tag{25}$$

If we denote by $\mathbf{D}(\sigma_i)$ the diagonal matrix of the standard deviations of the variates, the covariance and correlation matrices can be related as

$$\begin{aligned} \mathbf{P} &= \mathbf{D}\left(\frac{1}{\sigma_i}\right)\boldsymbol{\Sigma}\mathbf{D}\left(\frac{1}{\sigma_i}\right) \\ \boldsymbol{\Sigma} &= \mathbf{D}(\sigma_i)\mathbf{P}\mathbf{D}(\sigma_i) \end{aligned} \tag{26}$$

1.4 The Multivariate Normal Distribution

In the remainder of this book only the multivariate normal distribution will be used to describe the population out of which our samples of observation vectors will be drawn. There are two compelling reasons for this restriction:

1. A random vector which arose as the sum of a large number of independently and identically distributed random vectors will be distributed according to the multivariate normal distribution as the number of these fundamental source vectors increases without bound. That is, the usual *central-limit theorem*, which assures a normal distribution for variates which are summations of many independent random inputs, carries over directly to multidimensional inputs. This summation model appears to be a realistic one for many kinds of random phenomena encountered in the life and behavioral sciences.
2. Different models for the variate vectors might lead to rather different joint distributions of the elements whose mathematical complexity would prevent the development of the sampling distributions of the usual test statistics and estimates. Such distributions would have to be provided for each model's fundamental population. However, it seems likely that with the exception of rather pathological cases, the multivariate central-limit theorem would guarantee that the large-sample distributions of test statistics would lead us to similar conclusions about the state of nature.

Now let us develop the multivariate normal density function. Recall that the density of a normally distributed random variable X is

$$\phi(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \qquad -\infty < x < \infty \tag{1}$$

The joint density of the independent normal variates is thus

$$\phi(x_1, \ldots, x_p) = \frac{1}{(2\pi)^{p/2}\sigma_1 \cdots \sigma_p} \exp\left[-\frac{1}{2}\sum_{i=1}^{p}\left(\frac{x_i - \mu_i}{\sigma_i}\right)^2\right] \tag{2}$$

If we write $\mathbf{x}' = [x_1, \ldots, x_p]$, $\mu' = [\mu_1, \ldots, \mu_p]$, and

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \cdots & \sigma_p^2 \end{bmatrix}$$

the joint density can be given as

$$\phi(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right] \tag{3}$$

In this representation we see immediately that x has been replaced by a vector variate, μ is now a vector of means, and σ^2 has been generalized to a diagonal matrix. The squared term of the univariate density exponent is now a quadratic form in the deviations of the variates from their means, and the square root of the determinant of $\boldsymbol{\Sigma}$ has assumed the role of the univariate scale factor σ.

The general p-dimensional normal density function is obtained by permitting $\boldsymbol{\Sigma}$ in (3) to be *any* $p \times p$ *symmetric positive definite* matrix. Then $\phi(\mathbf{x})$ is positive for all finite $\mathbf{x}$, and

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi(\mathbf{x}) dx_1 \cdots dx_p = 1$$

for all $\boldsymbol{\mu}$, so that $\phi(\mathbf{x})$ is indeed a density function. The ith element of $\boldsymbol{\mu}$ is still the mean of x_i, the ith diagonal element of the more general matrix $\boldsymbol{\Sigma}$ is still the ith variance, and now the ijth element σ_{ij} of $\boldsymbol{\Sigma}$ can be shown to be the covariance of the ith and jth components of $\mathbf{x}$. We see immediately that if all $p(p-1)/2$ covariances are zero, the p components of $\mathbf{x}$ are independently distributed.

The case of $p = 2$ is especially important in statistical theory. Here

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

and the joint density is

$$\phi(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2}\left[\frac{1}{1-\rho^2}\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1 - \mu_1}{\sigma_1}\right)\left(\frac{x_2 - \mu_2}{\sigma_2}\right) + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2\right]\right\} \tag{4}$$