WARNING

CONCERNING COPYRIGHT RESTRICTIONS

The Copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or reproduction.

One of three specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research.

If electronic transmission of reserve material is used for purposes in excess of what constitutes “fair use”, that user may be liable for copyright infringement.

This policy is in effect for the following document:

Morrison, Donald F.
Chapter 6: The Structure of Multivariate Observations I, Principal Components / from Multivariate Statistical Methods
Belmont, CA : Thomson/Brooks/Cole, 2005, 4th ed. pp. 264-316.

NO FURTHER TRANSMISSION OR DISTRIBUTION OF THIS MATERIAL IS PERMITTED

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

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

ISBN 0-534-38778-0

THIS BOOK PRINTED ON ACID-FREE RECYCLED PAPER

Chapter 6

The Structure of Multivariate Observations: I. Principal Components

6.1 Introduction

Earlier we discussed the use of partial, multiple, and canonical correlation for analyzing the dependence structure of a multinormal population. The proper use of those methods required that certain roles be assigned to some of the responses. For a partial correlation analysis it is necessary to decide which variables are to be correlated and which of the remaining responses must be held constant. Multiple correlation demands that one response be dependent upon some or all of the remaining variates. Similarly, for a canonical-correlation analysis the responses must be collected into two or more sets. All these choices depend upon the nature of the responses and other information external to the mere values of their correlations. The conclusions we may draw about the dependence structure will in turn depend upon those choices. Furthermore, if the analyses are repeated for different choices of the dependent or constant variates, the successive findings will hardly be independent or contain mutually exclusive bits of information about the structure.

It would seem clear that a new class of techniques will be required for picking apart the dependence structure when the responses are symmetric in nature or no *a priori* patterns of causality are available. Those methods fall under the general heading of *factor analysis*, for by them one attempts to descry those hidden factors which have generated the dependence or variation in the responses. That is, the observable, or *manifest*, variates are represented as functions of a smaller number of *latent* factor variates. The mathematical form of the functions must be one which will generate the covariances or correlations among the responses. If that form is simple, and if the latent variates are few in number, a more parsimonious description of the dependence structure can be obtained. Now for simplicity linear functions are difficult to surpass, and in the two principal techniques of these chapters we shall usually think of the responses and their observations as linear compounds of the latent variates. The analysis of the dependence structure amounts to the statistical estimation of the coefficients of the functions.

We shall begin our study by developing in this chapter the Hotelling principal-component technique. That methodology originated with K. Pearson (1901) as a means of fitting planes by orthogonal least squares, but was later proposed by Hotelling (1933, 1936a) for the particular purpose of analyzing correlation structures. We shall initially define the principal components of a multivariate sample statistically and algebraically and then in terms of the geometry of the scatter swarm of the observations. Some numerical methods for extracting components will be treated, and the problem of interpreting component coefficients will be illustrated by some examples from biology and cognitive psychology, and by some special patterned correlation matrices. Some results of Anderson and Lawley on the sampling properties of principal components will be discussed and illustrated in the penultimate section.

6.2 The Principal Components of Multivariate Observations

Suppose that the random variables $X_1, \ldots, X_p$ of interest have a certain multivariate distribution with mean vector μ and covariance matrix Σ. We assume, of course, that the elements of μ and Σ are finite. The rank of Σ is $r \leq p$, and the q largest characteristic roots

$$\lambda_1 > \cdots > \lambda_q$$

of Σ are all distinct. For the present we shall not require a multinormal distribution of the X_i.

From this population a sample of N independent observation vectors has been drawn. The observations can be written as the usual $N \times p$ data matrix

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

Here a cautionary note on the ranks of Σ and $\mathbf{X}$ is in order. Mathematically, those matrices need not be of full rank p, nor need Σ contain more than one distinct characteristic root. However, the exigency of simplicity in our description of the latent structure of the X_i calls for a data matrix of full rank. We do not wish to confound the problem by including as responses total scores, weighted averages suggested by earlier studies, or other linear compounds which will reduce the rank of $\mathbf{X}$ and obscure whatever latent structure may be present.

The estimator of Σ will be the usual sample covariance matrix $\mathbf{S}$ defined by (9) of Section 1.6. The information we shall need for our principal-component analysis will be contained in $\mathbf{S}$. However, it will be necessary to make a choice of measures of dependence: Should we work with the variances and covariances of the observations, and carry out our analyses in the original units of the responses, or would a more accurate picture of the dependence pattern be obtained if each x_{ij}

were transformed to a standard score

(2) $$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$$

and the correlation matrix employed? The components obtained from **S** and **R** are in general not the same, nor is it possible to pass from one solution to the other by a simple scaling of the coefficients. Most applications of the technique have involved the correlation matrix, as if in keeping with the usage established by factor analysts. If the responses are in widely different units (age in years, weight in kilograms, and biochemical excretions in a variety of units, to cite one plausible case), linear compounds of the original quantities would have little meaning and the standardized variates and correlation matrix should be employed. Conversely, if the responses are reasonably commensurable, the covariance form has a greater statistical appeal, for as we shall presently see, the ith principal component is that linear compound of the responses which explains the ith largest portion of the total response variance, and maximization of such total variance of standard scores has a rather artificial quality (Anderson, 1963a, p. 139). Furthermore, as Anderson has shown, the sampling theory of components extracted from correlation matrices is exceedingly more complex than that of covariance-matrix components.

The first principal component of the observations **X** is that linear compound

(3) $$\begin{aligned} Y_1 &= a_{11}X_1 + \cdots + a_{p1}X_p \\ &= \mathbf{a}_1'\mathbf{x} \end{aligned}$$

of the responses whose sample variance

(4) $$\begin{aligned} s_{Y_1}^2 &= \sum_{i=1}^{p}\sum_{j=1}^{p} a_{i1}a_{j1}s_{ij} \\ &= \mathbf{a}_1'\mathbf{S}\mathbf{a}_1 \end{aligned}$$

is greatest for all coefficient vectors normalized so that $\mathbf{a}_1{}'\mathbf{a}_1 = 1$. To determine the coefficients we introduce the normalization constraint by means of the Lagrange multiplier l_1 and differentiate with respect to $\mathbf{a}_1$:

(5) $$\begin{aligned} \frac{\partial}{\partial \mathbf{a}_1}\left[s_{Y_1}^2 + l_1(1 - \mathbf{a}_1'\mathbf{a}_1)\right] &= \frac{\partial}{\partial \mathbf{a}_1}[\mathbf{a}_1'\mathbf{S}\mathbf{a}_1 + l_1(1 - \mathbf{a}_1'\mathbf{a}_1)] \\ &= 2(\mathbf{S} - l_1\mathbf{I})\mathbf{a}_1 \end{aligned}$$

The coefficients must satisfy the p simultaneous linear equations

(6) $$(\mathbf{S} - l_1\mathbf{I})\mathbf{a}_1 = \mathbf{0}$$

If the solution to these equations is to be other than the null vector, the value of l_1 must be chosen so that

(7) $$|\mathbf{S} - l_1\mathbf{I}| = 0$$

l_1 is thus a characteristic root of the covariance matrix, and $\mathbf{a}_1$ is its associated characteristic vector. To determine which of the p roots should be used, premultiply the system of equations (6) by $\mathbf{a}_1'$. Since $\mathbf{a}_1'\mathbf{a}_1 = 1$, it follows that

$$l_1 = \mathbf{a}_1'\mathbf{S}\mathbf{a}_1 = s_{Y_1}^2 \tag{8}$$

But the coefficient vector was chosen to maximize this variance, and l_1 must be the *greatest* characteristic root of $\mathbf{S}$. Let us summarize these results in this form:

Definition 6.1. The first principal component of the complex of sample values of the responses $X_1, \ldots, X_p$ is the linear compound

$$Y_1 = a_{11}X_1 + \cdots + a_{p1}X_p \tag{9}$$

whose coefficients a_{i1} are the elements of the characteristic vector associated with the greatest characteristic root l_1 of the sample covariance matrix of the responses. The a_{i1} are unique up to multiplication by a scale factor, and if they are scaled so that $\mathbf{a}_1'\mathbf{a}_1 = 1$, the characteristic root l_1 is interpretable as the sample variance of Y_1.

But what is the utility of this artificial variate constructed from the original responses? In the extreme case of $\mathbf{X}$ of rank one the first principal component would explain all the variation in the multivariate system. In the more usual case of the data matrix of full rank the importance and usefulness of the component would be measured by the proportion of the total variance attributable to it. If 87% of the variation in a system of six responses could be accounted for by a simple weighted average of the response values, it would appear that almost all the variation could be expressed along a single continuum rather than in six-dimensional space. Not only would this appeal to our sense of parsimony, but the coefficients of the six responses would indicate the relative importance of each original variate in the new derived component.

The second principal component is that linear compound

$$Y_2 = a_{12}X_1 + \cdots + a_{p2}X_p \tag{10}$$

whose coefficients have been chosen, subject to the constraints

$$\begin{aligned} \mathbf{a}_2'\mathbf{a}_2 &= 1 \\ \mathbf{a}_1'\mathbf{a}_2 &= 0 \end{aligned} \tag{11}$$

so that the variance of Y_2 is a maximum. The first constraint is merely a scaling to assure the uniqueness of the coefficients, while the second requires that $\mathbf{a}_1$ and $\mathbf{a}_2$ be orthogonal. The immediate consequence of the orthogonality is that the variances of the successive components sum to the total variance of the responses. The geometric implication will become clear in the next section. The coefficients of the second component are found by introducing the constraints (11) by the Lagrange multipliers l_2 and μ and differentiating with respect to $\mathbf{a}_2$:

$$\frac{\partial}{\partial \mathbf{a}_2}\left[\mathbf{a}_2'\mathbf{S}\mathbf{a}_2 + l_2\left(1 - \mathbf{a}_2'\mathbf{a}_2\right) + \mu\mathbf{a}_1'\mathbf{a}_2\right] = 2(\mathbf{S} - l_2\mathbf{I})\mathbf{a}_2 + \mu\mathbf{a}_1 \tag{12}$$

If the right-hand side is set equal to $\mathbf{0}$ and premultiplied by $\mathbf{a}_1'$, it follows from the normalization and orthogonality conditions that

$$2\mathbf{a}_1'\mathbf{S}\mathbf{a}_2 + \mu = 0 \tag{13}$$

Similar premultiplication of the equations (6) by $\mathbf{a}_2'$ implies that

$$\mathbf{a}_1'\mathbf{S}\mathbf{a}_2 = 0 \tag{14}$$

and hence $\mu = 0$. The second vector must satisfy

$$(\mathbf{S} - l_2\mathbf{I})\mathbf{a}_2 = \mathbf{0} \tag{15}$$

and it follows that the coefficients of the second component are thus the elements of the characteristic vector corresponding to the second greatest characteristic root. The remaining principal components are found in their turn from the other characteristic vectors. Let us summarize the process in this formal definition:

Definition 6.2. The jth principal component of the sample of p-variate observations is the linear compound

$$Y_j = a_{1j}X_1 + \cdots + a_{pj}X_p \tag{16}$$

whose coefficients are the elements of the characteristic vector of the sample covariance matrix $\mathbf{S}$ *corresponding to the jth largest characteristic root* l_j. *If* $l_i \neq l_j$, *the coefficients of the ith and jth components are necessarily orthogonal; if* $l_i = l_j$, *the elements can be chosen to be orthogonal, although an infinity of such orthogonal vectors exists. The sample variance of the jth component is* l_j, *and the total system variance is thus*

$$l_1 + \cdots + l_p = \text{tr}\,\mathbf{S} \tag{17}$$

The importance of the jth component in a more parsimonious description of the system is measured by

$$\frac{l_j}{\text{tr}\,\mathbf{S}} \tag{18}$$

The algebraic sign and magnitude of a_{ij} indicate the direction and importance of the contribution of the ith response to the jth component. A more precise and widely used statistical interpretation is also available. The sample covariances of the responses with the jth component are given by the column vector

$$\mathbf{S}\mathbf{a}_j \tag{19}$$

By the definition $(\mathbf{S} - l_j\mathbf{I})\mathbf{a}_j = \mathbf{0}$ of $\mathbf{a}_j$,

$$\mathbf{S}\mathbf{a}_j = l_j\mathbf{a}_j \tag{20}$$

and the covariance of the ith response with Y_j is merely $l_j a_{ij}$. If we divide by the component and response standard deviations, it follows that

$$\frac{a_{ij}\sqrt{l_j}}{s_i} \tag{21}$$

is the product-moment correlation of the ith response and the jth component. If the components have been extracted from the correlation matrix, the correlations of the responses with the jth component are given by the vector $\sqrt{l_j}\mathbf{a}_j$. In presenting components in this chapter we shall usually adopt that form of weight.

The vectors $\sqrt{l_j}\mathbf{a}_j$ bear an important relation to the correlation or covariance matrix from which they were extracted. The diagonalization of a square matrix implies that every real symmetric matrix $\mathbf{S}$ can be written as

$$\mathbf{S} = \mathbf{P}\mathbf{D}(l_i)\mathbf{P}'$$

where $\mathbf{P}$ is an orthogonal matrix and $\mathbf{D}(l_i)$ is the diagonal matrix of the characteristic roots of $\mathbf{S}$. If we take as columns of $\mathbf{P}$ the characteristic vectors of $\mathbf{S}$, it follows that

$$\mathbf{S} = \mathbf{P}\mathbf{D}(\sqrt{l_i})D(\sqrt{l_i})\mathbf{P}' \tag{22}$$

Let

$$\mathbf{L} = \mathbf{P}\mathbf{D}(\sqrt{l_i})$$

Then the columns of $\mathbf{L}$ *reproduce* $\mathbf{S}$ by the relation

$$\begin{aligned} \mathbf{S} &= l_1\mathbf{a}_1\mathbf{a}_1' + \cdots + l_r\mathbf{a}_r\mathbf{a}_r' \\ &= \mathbf{L}\mathbf{L}' \end{aligned} \tag{23}$$

The rank r of $\mathbf{S}$ may be less than p. As successive components are extracted from $\mathbf{S}$, the matrices $l_i\mathbf{a}_i\mathbf{a}_i'$ can be formed and their running sum compared with $\mathbf{S}$ to determine how well that matrix is being generated by a smaller number of variates.

By the relation (23) principal-component analysis is equivalent to a factorization of $\mathbf{S}$ into the product of a matrix $\mathbf{L}$ and its transpose. As we shall see in the next chapter, this is also the purpose of factor analysis wherein "factorization" of a matrix has precisely that algebraic meaning. *However,* in component analysis this factorization is unique up to the coefficient signs, for the component coefficients have been chosen to partition the total variance orthogonally into successively smaller portions, and if the portions are distinct, only one set of coefficient vectors will accomplish this purpose. This uniqueness of component coefficients is frequently overlooked by some investigators, who subject every component matrix to a series of postmultiplications by orthogonal matrices to see which transformed set of weights has the simplest subject-matter interpretation. While the ability of the vectors to generate the original matrix $\mathbf{S}$ is unimpaired, their components no longer have the maximum-variance property.

If the components have been extracted from the correlation matrix rather than $\mathbf{S}$, the sum of the characteristic roots will be

$$\text{tr}\,\mathbf{R} = p \tag{24}$$

and the proportion of the total "variance" in the scatter of dimensionless standard scores attributable to the jth component will be l_j/p. The sum of the squared correlations $a_{ij}\sqrt{l_j}$ of the responses on that component will of course be the component variance l_j.

If the first r components explain a large amount of the total sample variance, they may be evaluated for each subject or sampling unit and used in later analyses in place of the original responses. For components extracted from the covariance matrix the component scores of the ith subject are

$$y_{i1} = \mathbf{a}_1'(\mathbf{x}_i - \bar{\mathbf{x}}), \quad \ldots, \quad y_{ir} = \mathbf{a}_r'(\mathbf{x}_i - \bar{\mathbf{x}}) \tag{25}$$

where $\mathbf{x}_i$ is the ith observation vector and $\bar{\mathbf{x}}$ is the sample mean vector. The scores can be written as the $N \times r$ matrix

$$\mathbf{Y} = \left(\mathbf{I} - \frac{1}{N}\mathbf{E}\right)\mathbf{XA} \tag{26}$$

where $\mathbf{X}$ is the data matrix (1), $\mathbf{E}$ is the $N \times N$ matrix of ones in every position, and $\mathbf{A}$ is the $p \times r$ matrix whose columns are the first r characteristic vectors. Had the $\mathbf{a}_i$ been extracted from the correlation matrix, the scores would be computed from the standardized observations. Thus, the component values of the ith subject would be

$$y_{i1} = \mathbf{a}_1'\mathbf{z}_i, \quad \ldots, \quad y_{ir} = \mathbf{a}_r'\mathbf{z}_i \tag{27}$$

where $\mathbf{z}_i$ is the vector of standard scores with jth element given by equation (2).

If the ith and jth principal components correspond to distinct characteristic roots, their sample values will be uncorrelated. We may verify this by premultiplying the matrix equation (20) by the ith characteristic vector:

$$\mathbf{a}_i'\mathbf{S}\mathbf{a}_j = l_j\mathbf{a}_i'\mathbf{a}_j \tag{28}$$

But the left-hand side is merely the sample covariance of the component values $y_{hi} = \mathbf{a}_i'\mathbf{x}_h$, $y_{hj} = \mathbf{a}_j'\mathbf{x}_h$, and if $l_i \neq l_j$, it follows from the orthogonality of the vectors that

$$\mathbf{a}_i'\mathbf{S}\mathbf{a}_j = 0$$

How Many Components Should We Use?

To a certain extent principal-components analysis is similar to the selection of predictor variables in multiple regression: We wish to explain as much of the total variation in the data as possible with a minimum of components. If the first two or three components from a complex of several variables account for 80% or more of the total variance we have probably attained an appreciable reduction in dimensionality, and should stop with the interpretation of those components. If that proportion cannot be explained by the first four or five components it is usually fruitless to persist in attempting to name the later vectors. Conversely, if the first several components explain only a moderate proportion of the total variance the stopping rule is less clear. Usually we discard the later components whose coefficients have little or no substantive meaning. We might also ignore components with nearly equal variance, since their coefficients are poorly defined.

One approach consists of comparing the successive component variances and discarding the later components whose characteristic roots are declining in size more slowly. The *scree plot* proposed by Cattell (1966) is a graphical way of determining where the point of "decreasing returns" of explained variance occurs. The successive characteristic roots are simply plotted against their ordinal numbers. If the descent of the plot slows abruptly, the subsequent components from the "elbow" of the plot onward are dropped from consideration. If the plot does not have the "elbow" shape it is unlikely that a small number of components will suffice to explain the variation in the data. Scree plots are available as a graphics option in the principal-components procedure of the MINITAB (1997) Release 12 Statistical Software system.

Example 6.1. We shall illustrate a scree plot with the principal components of an 11×11 covariance matrix of course evaluation scores given by $N = 16$ students in a graduate statistics class. The evaluation scales ran from 1 ("poor" or "not at all") to 5 ("excellent," "strongly," or "difficult"). The data and the component coefficients will be omitted to save space. This scree plot of Figure 6.1 of the eigenvalues, or characteristic roots, was produced by MINITAB.

The decline in the size of the characteristic roots appears to begin with the third component. The first two components account for 76.0% of the total variance, and probably are the only ones that should be retained. The last four components only explain 2.2% of the variance, and suggest that the real dimensionality of the data complex is much smaller than eleven.

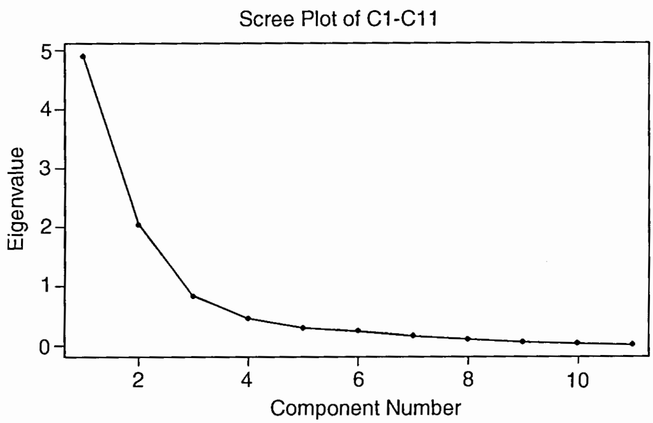

FIGURE 6.1 Scree plot of the course evaluation component variances

McCabe (1984) has summarized some optimal properties of principal components, and has proposed an alternative notion of "principal variables" for explaining variation in a multivariate sample. The principal variables avoid the complications of interpreting components that are linear compounds of the observed variables.

Example 6.2. Before turning to the geometric interpretation of principal components and a computing algorithm let us fix the ideas of this section with an example from biometry. Jolicoeur and Mosimann (1960) have investigated the principal components of carapace length, width, and height of painted turtles in an effort to give meanings to the concepts of "size" and "shape". The covariance matrix of the lengths, widths, and heights in millimeters of the carapaces of 24 female turtles was

$$\mathbf{S} = \begin{bmatrix} 451.39 & 271.17 & 168.70 \\ & 171.73 & 103.29 \\ & & 66.65 \end{bmatrix}$$

The coefficients and variances of the three components extracted from this matrix are summarized in Table 6.1.

The first principal component accounts for nearly all the variance in the three dimensions. It is the new weighted mean of the carapace measurements

$$Y_1 = 0.81(\text{length}) + 0.50(\text{width}) + 0.31(\text{height})$$

The size of the turtle shells could be characterized by this variable with little loss of information. Had the dimensions been expressed in logarithms of units, Y_1 would indeed be the logarithm of the volume of a box whose sides were powers of the actual carapace dimensions. Jolicoeur and Mosimann call the second and third components measures of carapace "shape", for they appear to be comparisons of length versus width and height, and height versus length and width, respectively. We shall consider a test of the uniqueness of variances of these components in Section 6.6. The component correlation coefficients are given in Table 6.2.

Component 1 appears to be almost equally correlated with the three dimensions. Components 2 and 3 are correlated with the width and height dimensions, but then only to a negligible degree.

TABLE 6.1
Carapace component coefficients

	Component		
Dimension	**1**	**2**	**3**
Length	0.8126	−0.5454	−0.2054
Width	0.4955	0.8321	−0.2491
Height	0.3068	0.1006	0.9465
Variance	680.40	6.50	2.86
Percentage of total variance	98.64	0.94	0.41

TABLE 6.2
Correlations of carapace dimensions and components

	Component		
Dimension	**1**	**2**	**3**
Length	1.00	−0.07	−0.02
Width	0.99	0.16	−0.03
Height	0.98	0.03	0.20

Computation of the Characteristic Roots and Vectors

We shall assume in this chapter that a subroutine or program is available for extracting all of the characteristic roots and vectors of square symmetric matrices of reasonable dimensions. BMDP Program P4M (BMDP, 1985) will give the principal components of covariance or correlation matrices in loading, or variable and component correlation, form. The components are computed as the initial stage of the *factor analyses* we shall describe in the next chapter. Similarly, the SAS (1979) statistical package will give the principal components of a data set as an option in the PROC FACTOR procedure. In both systems the principal component method is the "default" option when no method of factor extraction is specified. The programming language APL2 (Gilman and Rose, 1984) contains a primitive function which, when applied to a square matrix, produces a new matrix whose first row contains the characteristic roots of the original matrix, and whose successive rows are the characteristic vectors corresponding to the roots. STSC (1989) APL PLUS contains functions for extracting characteristic roots and vectors ("eigenvalues" and "eigenvectors" in its usage) from square matrices. Such interactive statistical packages as the STSC STATGRAPHICS™ (1989, Chapter 25) and MINITAB (Ryan *et al.*, 1980) contain procedures for extracting the principal components of covariance and correlation matrices.

Numerical analysis texts contain many methods for extracting characteristic roots and vectors. We shall only mention an iterative scheme proposed by Hotelling (1936a) which can be implemented very well with a language such as APL, and which will usually serve admirably when a formal package or program is not available. Select an initial vector $\mathbf{a}_0$, for example, $\mathbf{a}_0' = [1, \ldots, 1]$, compute the sequence of column vectors

$$\begin{aligned} \mathbf{a}^{(1)} &= \mathbf{S}\mathbf{a}_0 \\ \mathbf{a}^{(2)} &= \mathbf{S}^2\mathbf{a}_0 \\ \mathbf{a}^{(3)} &= \mathbf{S}^4\mathbf{a}_0 \\ \mathbf{a}^{(4)} &= \mathbf{S}^8\mathbf{a}_0 \\ &\vdots \end{aligned} \tag{29}$$

and normalize each $\mathbf{a}^{(j)}$ to unit length. The sequence of normalized vectors should converge to the vector $\mathbf{a}_1$ corresponding to the largest characteristic root, which can be computed as

$$l_1 = \mathbf{a}_1'\mathbf{S}\mathbf{a}_1 \tag{30}$$

by expression (8), Section 6.2. If the initial vector $\mathbf{a}_0$ is not too dissimilar from $\mathbf{a}_1$, convergence should occur in three or four iterations. If the covariances of $\mathbf{S}$ have mixed signs, or if the variances are of different magnitudes, the signs and sizes of the elements of $\mathbf{a}_0$ should be changed to reflect the likely pattern of the components in $\mathbf{a}_1$. The second characteristic vector and root can be found by applying the iterative process to the "deflated" matrix

$$\mathbf{S} - l_1\mathbf{a}_1\mathbf{a}_1' \tag{31}$$

where of course $\mathbf{a}_1'\mathbf{a}_1 = 1$.

Example 6.3. We shall extract the first principal component from the matrix $\mathbf{S}$ in Example 6.2 by the powering algorithm. The initial vector is $\mathbf{a}_0' = [1,1,1]$. The normalized vectors from the first through fourth iterations are shown in the table:

	Iteration			
Variable	**1**	**2**	**3**	**4**
Length	0.8111267612	0.8126297093	0.8126427616	0.8126427627
Width	0.4970820251	0.4955115187	0.4954946280	0.4954946263
Height	0.3081928578	0.3067593362	0.3067520428	0.3067520425

The vectors converged to four-place accuracy in only two iterations.

6.3 The Geometrical Meaning of Principal Components

We have introduced principal components analytically as those linear combinations of the responses which explain progressively smaller portions of the total sample variance. We shall now discuss the geometrical interpretation of components as the variates corresponding to the principal axes of the scatter of the observations in space. Imagine that a sample of N trivariate observations has the scatter plot shown in Figure 6.2, where the origin of the response axes has been taken at the sample means. The swarm of points seems to have a generally ellipsoidal shape, with a major axis Y_1 and less well-defined minor axes Y_2 and Y_3. Let us confine our attention for the moment to the major axis and denote its angles with the original response axes as α_1, α_2, and α_3. If Y_1 passes through the sample mean point, its orientation is completely determined by the direction cosines

$$a_{11} = \cos\alpha_1 \quad a_{21} = \cos\alpha_2 \quad a_{31} = \cos\alpha_3 \tag{1}$$

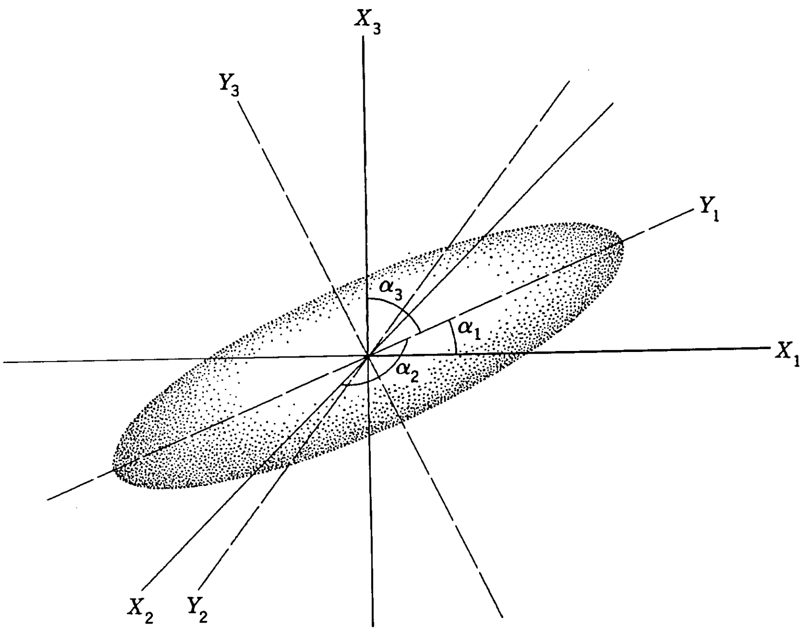

FIGURE 6.2 Principal axes of trivariate observations

where $a_{11}^2 + a_{21}^2 + a_{31}^2 = 1$. It is known from analytic geometry (Eisenhart, 1960; Somerville, 1958) that the value of the observation $[x_{i1},\ x_{i2},\ x_{i3}]$ on the new coordinate axis Y_1 will be

(2) $$y_{i1} = a_{11}(x_{i1} - \bar{x}_1) + a_{21}(x_{i2} - \bar{x}_2) + a_{31}(x_{i3} - \bar{x}_3)$$

Note that the mean of the Y_1 variate is

(3) $$\begin{aligned} \bar{y}_1 &= \frac{1}{N}\sum_{i=1}^{N} y_{i1} \\ &= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{3} a_{j1}(x_{ij} - \bar{x}_j) \\ &= 0 \end{aligned}$$

Let us explicate the notion of the *major* axis of a discrete swarm of points by defining that axis as passing through the direction of *maximum variance* in the points. In the present case of three responses that variance is

(4) $$\frac{1}{N-1}\sum_{i=1}^{N} y_{i1}^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left[\sum_{j=1}^{3} a_{j1}(x_{ij} - \bar{x}_j)\right]^2$$

and the angles of Y_1 would be found by differentiating this expression with respect to the a_{j1} (with suitable provision for the constraint) and solving for the a_{j1} which make the derivatives zero. The solution would be the characteristic vector of the greatest root of the sample covariance matrix of the x_{ij}, and Y_1 would be the continuum of the first *principal component* of the system.

Let us prove this statement for the general case of p responses. Write the direction cosines of the first principal-component axis as $\mathbf{a}_1' = [a_{11}, \ldots, a_{p1}]$, where the constraint

(5)
$$\mathbf{a}_1'\mathbf{a}_1 = 1$$

must always be satisfied. The variance of the projections on the Y_1 axis is

(6)
$$\begin{aligned}
s_{Y_1}^2 &= \frac{1}{N-1}\sum_{i=1}^{N} y_{i1}^2 \\
&= \frac{1}{N-1}\sum_{i=1}^{N}\left[\sum_{j=1}^{p} a_{j1}(x_{ij} - \bar{x}_j)\right]^2 \\
&= \frac{1}{N-1}\sum_{i=1}^{N}[(\mathbf{x}_i - \bar{\mathbf{x}})'\mathbf{a}_1]^2 \\
&= \frac{1}{N-1}\sum_{i=1}^{N}\mathbf{a}_1'(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'\mathbf{a}_1 \\
&= \mathbf{a}_1'\mathbf{S}\mathbf{a}_1
\end{aligned}$$

Introduce the constraint (5) by the Lagrange multiplier l_1. Then the maximand

(7)
$$\mathbf{a}_1'\mathbf{S}\mathbf{a}_1 + l_1(1 - \mathbf{a}_1'\mathbf{a}_1)$$

is precisely the same as that of equation (5) in Section 6.2. The direction cosines of the first principal axis are the elements of the first characteristic vector of $\mathbf{S}$, and the maximized variance is the greatest characteristic root.

The remaining characteristic roots and vectors of $\mathbf{S}$ determine the lengths and orientations of the second and higher component axes. If two successive roots l_i and l_{i+1} are equal, the scatter configuration has no unique major axis in the plane of the axes of the roots and the appearance of the points is more circular than elliptical. Such dispersion is called *isotropic* or *spherical* in those dimensions with equal l_i. These results may be summarized in this geometrical definition of principal components:

Definition 6.3. The principal components of the sample of N p-dimensional observations are the new variates specified by the axes of a rigid rotation of the original response coordinate system into an orientation corresponding to the directions of maximum variance in the sample scatter configuration. The direction cosines of the new axes are the normalized characteristic vectors corresponding to the successively smaller characteristic roots of the sample covariance matrix. If two or more roots are equal, the directions of the associated axes are not unique and may be chosen in an infinity of orthogonal positions. If the components are instead computed from the correlation matrix, the same geometrical interpretation holds, although the response coordinate system is expressed in standard units of zero means and unit variances.

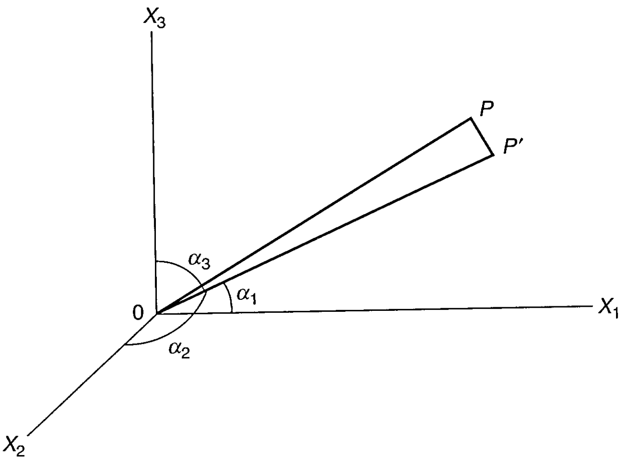

FIGURE 6.3 Observation point p and its projection p' on the first principal axis

A second property of component axes was first considered by K. Pearson (1901) and is implicit in the preceding geometrical derivation: The choice of the new coordinate axes is such that the sums of squared distances of each point to its projections on the successive axes are minimized. Figure 6.3 shows one projection and distance of an observation in three-dimensional space. In the general case the squared ith distance is

$$\begin{aligned}(P_i'P_i)^2 &= (OP_i)^2 - (OP_i')^2 \\ &= \sum_{j=1}^{p}(x_{ij}-\bar{x}_j)^2 - \left[\sum_{j=1}^{p} a_{j1}(x_{ij}-\bar{x}_j)\right]^2\end{aligned} \tag{8}$$

Minimization of the sum $\sum_{i=1}^{N}(P_i'P_i)^2$ of all squared distances to the new axis is equivalent to maximization of the second term of (8), which is of course proportional to the component variance (6). The orthogonal least-squares solution for the best-fitting line leads immediately to the first principal axis, and the other component axes would follow from the successive orthogonal least-squares solutions.

Jong and Kotz (1999) have given an algebraic derivation of principal components as the orthogonal linear functions that minimize the regression model error sum of squares.

If the first characteristic root is large, an excellent approximation to the coefficients of the first component can be obtained from the line connecting the two extreme points in the scatter configuration. Denote by $\mathbf{x}_{(1)}$ and $\mathbf{x}_{(N)}$ the observations separated by the greatest distance in the sample space, or those for which

$$d^2 = \sum_{j=1}^{p}(x_{(N)j} - x_{(1)j})^2 \tag{9}$$

is a maximum for all pairs of observation vectors. The direction cosines of the line are

$$a_j = \frac{x_{(N)j} - x_{(1)j}}{d} \tag{10}$$

where d is the positive square root of the distance (9). These cosines or simply the differences $x_{(N)j} - x_{(1)j}$ can be used as the initial vector $\mathbf{x}_0$ for the iterative process of Section 6.2.

Some indication of the usefulness of this approximation follows from its application to the carapace data of Example 6.2. The approximate direction cosines of the greatest distance are [0.807, 0.519, 0.295], and these agree with the true values to nearly two places.

6.4 The Interpretation of Principal Components

In this section we shall discuss some examples of principal-component solutions in an attempt to give subject-matter identities to the new variables.

Example 6.4. *Alternate Analyses of the Turtle Carapace Dimensions.* It is evident from the covariance matrix of Example 6.2 that the three shell dimensions have very unequal variances. To stabilize the variances the observations on each shell were transformed to common logarithms. The new means and covariance matrix had these values:

Mean log dimension			
Length	**Width**	**Height**	**Covariance matrix** $\times 10^3$
2.128	2.008	1.710	$\begin{bmatrix} 4.9810 & 3.8063 & 4.7740 \\ & 3.0680 & 3.7183 \\ & & 4.8264 \end{bmatrix}$

The characteristic roots of the matrix are 12.6372, 0.1386, and 0.0997. Since 98.15 percent of the total variance of the log-transformed observations is attributable to the first component and the scatter in the remaining two dimensions about that axis is nearly circular, it is reasonable to conclude that the new variables lie about a new dimension with direction cosines

$$\mathbf{a}_1' = [0.6235, 0.4860, 0.6124]$$

The new component variable is

$$\begin{aligned} Y_1 &= 0.62 \log(\text{length}) + 0.49 \log(\text{width}) + 0.61 \log(\text{height}) \\ &= \log(X_1^{0.62} X_2^{0.49} X_3^{0.61}) \end{aligned}$$

or the logarithm of a kind of carapace volume.

Alternatively the dimensions might have been scaled to standard scores with unit variances. Then the component analysis would be carried out upon the correlation matrix of

$$\begin{bmatrix} 1 & 0.974 & 0.973 \\ & 1 & 0.966 \\ & & 1 \end{bmatrix}$$

and as we might expect from such nearly equal correlations the direction cosine vector

$$\mathbf{a}_1' = [0.5783,\ 0.5770,\ 0.5768]$$

of the first principal axis of the standard scores is almost exactly that of the equiangular line in three-space. This component accounts for 98.0 percent of the "variance" of the standard scores. The remaining two dimensions account for nearly equal parts of the total variance, and we conclude as before that removal of the inequalities of the variances produces a scatter with a single long axis surrounded by fairly isotropic variation.

Example 6.5. *Wechsler Adult Intelligence Scale Subtest Scores.* The correlation coefficients of principal components extracted from the correlation matrix given by Birren and Morrison (1961) for the 11 WAIS subtest scores, age, and years of formal education of 933 white, native-born male and female participants in a community testing program are shown in Table 6.3. The first component appears to be a measure of *general intellectual performance*, for all WAIS subtests have nearly equal high positive correlations with this dimension. The coefficient of education is also positive and of exactly the order of size as the subtest coefficients. Age is negatively represented in this component, and with a smaller absolute weight than those of the other variables. More than half the total variance in the 13 original dimensions is accounted for by component 1.

Component 2 explained a much smaller percentage (10.90 %) of the total variance. This dimension may be thought of as an *experiential* or *age* factor, for that variable dominates the coefficients. In

TABLE 6.3
Correlation coefficients of the WAIS principal components

	Component			
	1	2	3	4
WAIS Subtests:				
Information	0.83	0.33	−0.04	−0.10
Comprehension	0.75	0.31	0.07	−0.17
Arithmetic	0.72	0.25	−0.08	0.35
Similarities	0.78	0.14	0.00	−0.21
Digit span	0.62	0.00	−0.38	0.58
Vocabulary	0.83	0.38	−0.03	−0.16
Digit symbol	0.72	−0.36	−0.26	−0.08
Picture completion	0.78	−0.10	−0.25	−0.01
Block design	0.72	−0.26	0.36	0.18
Picture arrangement	0.72	−0.23	0.04	−0.05
Object assembly	0.65	−0.30	0.47	0.13
Age	−0.34	0.80	0.26	0.18
Years of education	0.75	0.01	−0.30	−0.23
Latent root (variance)	6.69	1.42	0.80	0.71
Percentage of total variance	51.47	10.90	6.15	5.48
Cumulative percentage of variance	51.47	62.37	68.52	74.01

addition, the verbal subtests (1 to 6) have positive coefficients, while the performance tests (7 to 11) correlate negatively. In that sense component 2 is a bipolar dimension comparing one set of verbal or informational skills known to increase with advancing age to subtests measuring spatial-perceptual qualities and other cognitive abilities known to decrease with age.

The third and fourth components explain small and nearly equal proportions of the total variance. Component 3 is highly correlated with block design and object assembly, though negatively with digit span, and may be thought of as a *spatial imagery or perception* dimension. Component 4 appears to be a measure of *numerical facility.*

Example 6.6. *Another Biometric Application: Fowl-Bone Lengths.* Wright (1954) reported these correlations among six bone dimensions of $N = 276$ white Leghorn fowl:

$$\begin{array}{l} \text{Skull length} \\ \text{Skull breadth} \\ \text{Humerus} \\ \text{Ulna} \\ \text{Femur} \\ \text{Tibia} \end{array}
\begin{bmatrix} 1.000 & 0.584 & 0.615 & 0.601 & 0.570 & 0.600 \\ & 1.000 & 0.576 & 0.530 & 0.526 & 0.555 \\ & & 1.000 & 0.940 & 0.875 & 0.878 \\ & & & 1.000 & 0.877 & 0.886 \\ & & & & 1.000 & 0.924 \\ & & & & & 1.000 \end{bmatrix}$$

These anatomical measurements have a *hierarchical* pattern: the skull contributes two dimensions, while two wing bones (humerus and ulna) and two leg bones (femur and tibia) are also represented. This partitioning of the matrix is apparent in the values of the correlation coefficients. It is even more striking in the patterns of the component correlations obtained by Wright (1954) and reproduced here in Table 6.4.

As a practical problem in principal axes or data reduction one would probably stop with the first two or three components, for no parsimony is gained by merely making an orthogonal transformation to six new dimensions. However, it is a rare matrix that permits straightforward identifications of all its components:

Component	**Name**
1	General average of all bone dimensions
2	Comparison of skull size with the wing and leg lengths
3	Comparison of skull length and breadth; measure of skull shape
4	Comparison of wing and leg lengths
5	Comparison of femur and tibia
6	Comparison of humerus and ulna

We shall consider the case of hierarchically ordered variables in greater generality in the next section.

In an article on hypothesis tests for correlation-matrix principal components Schott (1991) has addressed the question of whether the first characteristic vector of the fowl-bone correlation matrix is the equiangular one $6^{-1/2}[1, 1, 1, 1, 1, 1]$. He conclusively rejected that hypothesis: The skull dimension coefficients are significantly smaller than those of the other bones. In a later article Schott (1997) has also considered some additional aspects of the fowl-bone component dimensionality and structure.

TABLE 6.4
Component correlations of the fowl-bone dimensions

Dimension	Component 1	2	3	4	5	6
Skull:						
Length	0.74	0.45	0.49	−0.02	0.01	0.00
Breadth	0.70	0.59	−0.41	0.00	0.00	−0.01
Wing:						
Humerus	0.95	−0.16	−0.03	0.22	0.05	0.16
Ulna	0.94	−0.21	0.01	0.20	−0.04	−0.17
Leg:						
Femur	0.93	−0.24	−0.04	−0.21	0.18	−0.03
Tibia	0.94	−0.19	−0.03	−0.20	−0.19	0.04
Characteristic root	4.568	0.714	0.412	0.173	0.076	0.057
Percentage of total variance	76.1	11.9	6.9	2.9	1.3	0.9

Example 6.7. Principal components were extracted from the WAIS scores of the no-senile-factor and senile-factor subjects given in Appendix B and described in Example 2.3, Section 2.4. The component coefficients and characteristic roots (rounded to save space) are shown in the table. With the exception of the subtest picture completion, the coefficients of the first principal component are approximately equal in the two groups. The first component explains a greater percentage of total variance in the senile-factor subjects. The other characteristic roots are successively of the same magnitudes, and suggest that a single-component analysis might be performed upon the within-groups covariance matrix of Example 2.3. The second component is largely a comparison of the arithmetic and similarities scores for the no-senile-factor sample, and a comparison of arithmetic and picture completion scores in the senile-factor group. The third and fourth components probably should be dismissed because of their small variances.

Subtest	No-senile-factor component 1	2	3	4	Senile-factor component 1	2	3	4
Information	0.61	−0.22	0.46	0.61	0.46	0.01	0.52	0.72
Similarities	0.60	−0.50	−0.32	−0.54	0.62	−0.17	−0.75	0.15
Arithmetic	0.51	0.79	−0.33	0.03	0.49	−0.49	0.40	−0.60
Picture completion	0.11	0.27	0.76	−0.58	0.40	0.85	0.07	−0.32
Characteristic root	26.25	6.26	3.93	1.91	41.61	7.98	3.84	1.31
Percent variance	68.45	16.31	10.25	4.99	76.02	14.58	7.01	3.39

The hypothesis of equal-covariance matrices for the two diagnostic populations is just tenable at the 0.05 level by the generalized likelihood-ratio statistic of Section 5.4, so we shall find the components of the within-groups covariance matrix. They are given in the table, along with the mean differences of the component scores for the two groups, and the squared critical ratios

$$t_j^2 = \frac{[\mathbf{a}'_j(\bar{\mathbf{x}}_{1j} - \bar{\mathbf{x}}_{2j})]^2 N_1 N_2}{l_j^2(N_1 + N_2)}$$

Because the component coefficients are subject to random variation the t_j should not be treated as t statistics, although their magnitudes give some sense of the separation of the two groups on the four principal-component dimensions. We note that the two-sample T^2 statistic of Example 2.3 is equal to

$$T^2 = t_1^2 + \cdots + t_4^2$$

The differences in mean subtest scores for the two diagnostic groups can be described as a level difference measured by the first component, and a measure of dissimilar mean profile shapes given by the third component. The mean differences of the second and fourth components can be ignored for their small t_j^2 ratios.

Within-groups principal components

Subtest	Component 1	2	3	4
Information	0.56	0.14	0.25	−0.77
Similarities	0.61	0.58	−0.05	0.54
Arithmetic	0.51	−0.67	−0.53	0.08
Picture completion	0.20	−0.45	0.81	0.33
Characteristic root	29.23	5.40	4.97	2.58
Percent variance	69.30	12.81	11.79	6.10
Group mean difference	6.94	−0.44	1.74	0.61
t_j^2	14.94	0.32	5.55	1.32

When the characteristic roots of several successive components are nearly equal the components are of course poorly defined. In this case Jolliffe (1989) has suggested that the component coefficients or correlations might be multiplied by an orthogonal matrix chosen to give more substantively meaningful and interpretable coefficients. This is the same process of *factor rotation* that we shall discuss in the next chapter. The principal-component sections of some computer packages contain rotation as an option. Jolliffe has given some examples of rotated components, as well as caveats for its use.

6.5 Some Patterned Matrices and Their Principal Components

The component structure of a covariance or correlation matrix can sometimes be approximated rather well by inspection of the elements and a knowledge of the characteristic roots and vectors of certain patterned matrices. We shall now give the components of two such special matrices, and a general upper bound on the greatest characteristic root of any square matrix.

Equicorrelation Matrix

Here the $p \times p$ covariance matrix is

(1)
$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \cdots & \cdots & \cdots & \cdots \\ \rho & \rho & \cdots & \rho \end{bmatrix}$$

We shall require that $0 < \rho \leq 1$. The greatest characteristic root of this matrix is

(2)
$$\lambda_1 = \sigma^2[1 + (p-1)\rho]$$

and its normalized characteristic vector is

(3)
$$\mathbf{a}_1' = \left[\frac{1}{\sqrt{p}}, \ldots, \frac{1}{\sqrt{p}}\right]$$

The first principal component

(4)
$$Y_1 = \frac{1}{\sqrt{p}} \sum_{j=1}^{p} X_j$$

is merely proportional to the mean of the original p responses; it accounts for $100[1 + (p-1)\rho]/p$ percent of the total variance in the set. The remaining $p - 1$ characteristic roots are all equal to

(5)
$$\sigma^2(1-\rho)$$

and their vectors are any of the $p - 1$ linearly independent solutions of the equation

(6)
$$\sigma^2\rho(a_{12} + a_{22} + \cdots + a_{p2}) = 0$$

Note that this equation amounts to the requirement that each of the $p - 1$ characteristic vectors be orthogonal to the first vector, or

(7)
$$\mathbf{a}_1'\mathbf{a}_2 = 0$$

Example 6.8. The largest characteristic root of the correlation matrix

$$\mathbf{P} = \begin{bmatrix} 1 & 0.6 & 0.6 & 0.6 \\ & 1 & 0.6 & 0.6 \\ & & 1 & 0.6 \\ & & & 1 \end{bmatrix}$$

is $\lambda_1 = 2.8$. Since the characteristic vector of that root is $\mathbf{a}' = [\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}]$, 70 percent of the total variance in such an idealized system would be explained by the first principal component

$$Y_1 = \frac{1}{2}(X_1 + X_2 + X_3 + X_4)$$

The remaining 30 percent of the variance is equally attributable to three new variates symmetrically distributed about the Y_1 axis.

The equal-variance and equal-covariance structure thus always contains a single principal component explaining a greater proportion of the total variance as the common correlation becomes large. This dimension has an *equiangular* orientation in the midst of the axes of the original variates. The other independent variates have isotropic variation in the remaining $p-1$ dimensions orthogonal to the major axis.

The Equipredictability Covariance Pattern

Bargmann (1957) has considered the properties of the patterned covariance matrix

$$\Sigma = \begin{bmatrix} \sigma^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ & \sigma^2 & \sigma_{14} & \sigma_{13} \\ & & \sigma^2 & \sigma_{12} \\ & & & \sigma^2 \end{bmatrix} \tag{8}$$

Such matrices have the property that the four multiple correlation coefficients of one variable with the remaining three are equal. The matrix has this principal-component structure:

Variance	Component direction cosines
$\sigma^2 + \sigma_{12} + \sigma_{13} + \sigma_{14}$	$[\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}]$
$\sigma^2 + \sigma_{12} - \sigma_{13} - \sigma_{14}$	$[\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}]$
$\sigma^2 - \sigma_{12} + \sigma_{13} - \sigma_{14}$	$[\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}]$
$\sigma^2 - \sigma_{12} - \sigma_{13} + \sigma_{14}$	$[\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}]$

For example, the correlation matrix

$$\begin{bmatrix} 1 & 0.7 & 0.6 & 0.4 \\ & 1 & 0.4 & 0.6 \\ & & 1 & 0.7 \\ & & & 1 \end{bmatrix}$$

has those characteristic vectors with roots 2.7, 0.7, 0.5, and 0.1, respectively. Using the fact that the characteristic vectors give the four orthogonal contrasts in the 2^2 factorial experimental design, Bock (1960) has established interesting connections between covariance structures and the model II analysis of variance.

An Upper Bound on the Greatest Characteristic Root of a Matrix

Brauer (1953) and other algebraists (*e.g.*, Marcus and Minc, 1964) have investigated the regions in which the characteristic roots of an arbitrary square matrix must lie. For the positive semidefinite covariance matrix Σ it can be shown that the maximum characteristic root cannot exceed the greatest of the row sums of absolute values of the elements of Σ:

$$\lambda_1 \leq \max\left(\sum_{j=1}^{p} |\sigma_{ij}|\right) \qquad i = 1, \ldots, p \tag{9}$$

If Σ is the equal-variance matrix (1) this bound is equal to the greatest root. As an example, the row sums of the fowl-bone dimensions correlation matrix of Example 6.6 are

3.970 3.771 4.884 4.834 4.772 4.843

The largest characteristic root must be less than 4.884, so that the first principal component can explain at most 81.4 percent of the total variance. Since this is a substantial proportion, the arithmetic labor might be begun with the assurance that the first component could possibly explain an appreciable amount of variance.

The use of this result for estimating λ_1 appears to be best reserved for correlation matrices, for the proximity of the greatest root to the upper bound diminishes as the matrix departs from the equal-variance, equal-covariance pattern. For example, the bound on the largest root 6.69 of the WAIS correlation matrix of Example 6.5 is 7.44. Even less precision is obtained if the inequality is applied to a covariance matrix with widely different diagonal elements, as for example, the carapace-dimension matrix of Example 6.2.

6.6 The Sampling Properties of Principal Components

Throughout the preceding sections we have treated principal-component analysis as a descriptive technique for studying the dependence or correlational structure of multivariate samples. Now we shall require that the sample has been drawn from a multinormal population whose covariance matrix has a specified covariance structure, and on the basis of that assumption we shall be able to state a number of large-sample distributional properties of the component coefficients and characteristic roots. In addition to providing knowledge of the stability of those quantities through their variances and covariance, these asymptotic distributions will also permit the construction of tests of hypotheses and confidence intervals for the population component structure. The initial derivation of these large-sample properties and tests was begun by Girshick (1936, 1939); subsequent extensions have been made by Anderson (1951a, 1963a), Bartlett (1954), and Lawley (1956, 1963). The results have been summarized by Anderson (1984) and Muirhead (1982) in their texts. Since the mathematics of that research is well above the level of this text, the derivations will be omitted.

Suppose that N independent observations have been taken on the p-dimensional random variable with distribution $N(\boldsymbol{\mu}, \Sigma)$. Σ has the *distinct* characteristic roots

$$\lambda_1 > \cdots > \lambda_p > 0 \tag{1}$$

with corresponding vectors

$$\boldsymbol{\alpha}_1, \ \ldots, \ \boldsymbol{\alpha}_p \tag{2}$$

The sample estimate of Σ is the matrix $\mathbf{S}$ with elements based upon n degrees of freedom, and if $\mathbf{S}$ has the usual single-sample form of Section 1.6, $n = N - 1$. Denote the roots and vectors of $\mathbf{S}$ by the Latin-letter counterparts

$$l_1 > \cdots > l_p > 0 \qquad (3)$$
$$\mathbf{a}_1, \ldots, \mathbf{a}_p$$

Both the population and sample characteristic vectors have the direction cosine form of unit lengths. Anderson (1984) has show that the maximum-likelihood estimators of the λ_i are given by the characteristic roots $(n/N)l_i$ of the maximum-likelihood estimator $\hat{\Sigma} = (n/N)\mathbf{S}$ of the population covariance matrix. The l_i are biased estimators of the λ_i. Lawley (1956) has computed the means and variances

$$E(l_i) \approx \lambda_i + \frac{\lambda_i}{n}\sum_{\substack{j=1\\j\neq i}}^{p}\frac{\lambda_j}{\lambda_i - \lambda_j} \qquad (4)$$
$$\operatorname{var}(l_i) \approx \frac{2\lambda_i^2}{n}\left[1 - \frac{1}{n}\sum_{\substack{j=1\\j\neq i}}^{p}\left(\frac{\lambda_j}{\lambda_i - \lambda_j}\right)^2\right]$$

to terms of order n^{-1}; Lawley (1956) and G. A. Anderson (1965) have found estimators with biases of the order of n^{-2}; they are

$$\hat{\lambda}_i = l_i\left(1 - \frac{1}{n}\sum_{\substack{j=1\\j\neq i}}^{p}\frac{l_j}{l_i - l_j}\right) \qquad (5)$$

Girshick and Anderson have demonstrated that the following results hold for distinct population roots as n becomes very large:

1. l_i is distributed independently of the elements of its associated vector $\boldsymbol{a}_i$.
2. $\sqrt{n}(l_i - \lambda_i)$ is normally distributed with mean zero and variance $2\lambda_i^2$ as n tends to infinity and is distributed independently of the other sample characteristic roots.
3. The elements of $\sqrt{n}(\mathbf{a}_i - \alpha_i)$ are distributed according to the multinormal distribution with null mean vector and covariance matrix

$$\lambda_i \sum_{\substack{h=1\\h\neq i}}^{p}\frac{\lambda_h}{(\lambda_h - \lambda_i)^2}\alpha_h\alpha_h' \qquad (6)$$

4. The covariance of the rth element of $\boldsymbol{a}_i$ and the sth element of $\boldsymbol{a}_j$ is

$$-\frac{\lambda_i \lambda_j \alpha_{si} \alpha_{rj}}{n(\lambda_i - \lambda_j)^2} \quad i \neq j \tag{7}$$

It is essential to bear in mind that these are *large*-sample distributional properties for the case of p *distinct* population characteristic roots. However, as Anderson has indicated, the result (3) requires that only λ_i be distinct from the other $p-1$ roots which may have any multiplicities.

The preceding general results can be used to construct tests of hypotheses and confidence intervals for the population roots and direction cosines. Perhaps a better idea of the sampling variation and dependence of component coefficients can be gained if the expressions are evaluated for a simple covariance matrix. If

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \cdots & \rho \\ \cdots & \cdots & \cdots \\ \rho & \cdots & 1 \end{bmatrix}$$

its roots and vectors have the forms given by (2) to (3) and (5) to (7) of Section 6.5. The variances and covariances of the direction cosines of the first principal component have the values

$$\begin{aligned} \operatorname{var}(a_{r1}) &= \frac{[1+(p-1)\rho](p-1)(1-\rho)}{np^3\rho^2} \\ \operatorname{cov}(a_{r1}, a_{s1}) &= -\frac{[1+(p-1)\rho](1-\rho)}{np^3\rho^2} \end{aligned} \tag{8}$$

The correlation of any two coefficients is thus $-1/(p-1)$. Some further numerical examples of the correlations among characteristic vector elements have been computed by Jackson and Hearne (1973).

Asymptotic confidence intervals for the ith covariance population root follow from result (2). For large n, $\sqrt{\frac{1}{2}n}(l_i - \lambda_i)/\lambda_i$ is a standard normal variate, and the probability is $1-\alpha$ that the inequality

$$\frac{n}{2\lambda_i^2}(l_i - \lambda_i)^2 \leq z_{1/2\alpha}^2 \tag{9}$$

holds, where $z_{1/2\alpha}$ denotes the upper 50α percentage point of the standard normal distribution. If the quadratic expression in λ_i is simplified, we have the more explicit $100(1-\alpha)$ percent asymptotic confidence interval:

$$\frac{l_i}{1 + z_{1/2\alpha}\sqrt{2/n}} \leq \lambda_i \leq \frac{l_i}{1 - z_{1/2\alpha}\sqrt{2/n}} \tag{10}$$

Of course n and α must always be such that the right-hand side of (10) is greater than zero.

Example 6.9. Let us find the 95% confidence interval for the smallest characteristic root of the carapace dimensions in Example 6.2. Here $n = 23$, $l_3 = 2.86$, $z_{1/2\alpha} = 1.96$, and the confidence interval is

$$1.81 \le \lambda_3 \le 6.78$$

This interval is of course in the original millimeters-squared units of the covariance matrix. The sample size of 24 observations is hardly of the "asymptotic" order of magnitude, and the true interval is probably considerably wider.

More General Inferences about the Dependence Structure

Anderson (1963a) has treated a test of the hypothesis

$$H_0: \quad \lambda_{q+1} = \cdots = \lambda_{q+r} \tag{11}$$

that r of the intermediate characteristic roots of Σ are equal. The q larger and $p-q-r$ smaller characteristic roots are unrestricted as to their values or multiplicities. The alternative hypothesis to H_0 is that some of the roots in the middle set are distinct. The likelihood-ratio criterion leads to the statistic

$$\chi^2 = -n \sum_j \ln l_j + nr \ln \frac{\sum_j l_j}{r} \tag{12}$$

where $n = N-1$ for the single-sample covariance matrix and the summations extend over the values $j = q+1, \ldots, q+r$. When the hypothesis (11) is true, the statistic has the chi-squared distribution with degrees of freedom $1/2r(r+1)-1$ for large n, and the hypothesis is of course rejected for large values of (12). An important special case of the hypothesis occurs when $q+r = p$ or when the variation in the last r dimensions is spherical. A slightly more conservative test of that hypothesis can be obtained by replacing n in (12) by $[n-(p-r)-(2r^2+r+2)/6r]$ (Bartlett, 1954).

Example 6.10. The hypothesis of equality of the second and third roots of the carapace data can be tested, albeit approximately because of the small sample size, by the statistic (12). Here $n = 23, q = 1, r = 2$, and

$$\begin{aligned}\chi^2 &= -23(\ln 6.50 + \ln 2.86) + 46 \ln \frac{6.50+2.86}{2} \\ &= 3.77\end{aligned}$$

$\chi^2_{0.05;2} = 5.99$, and at the $\alpha = 0.05$ level we would not reject the hypothesis that the normal population ellipsoid has minor axes of equal length. It is thus reasonable to think of the three shell dimensions as distributed about a single principal axis and two minor axes of isotropic variation. We note that acceptance of H_0 on the basis of the asymptotic chi-squared criterion also implies that the hypothesis would be accepted if the exact distribution of the statistic were known for the small sample with 23 degrees of freedom.

It is also possible to generalize the asymptotic confidence interval (10) to the case of multiple roots. If λ_i is of multiplicity r, the $100(1-\alpha)$ percent asymptotic confidence interval is

$$\frac{\bar{l}_i}{1 + z_{1/2\alpha}\sqrt{2/nr}} \le \lambda_i \le \frac{\bar{l}_i}{1 - z_{1/2\alpha}\sqrt{2/nr}} \tag{13}$$

where

$$\bar{l}_i = \frac{1}{r}(l_{q+1} + \cdots + l_{q+r}) \tag{14}$$

A one-sided $100(1 - 1/2\alpha)$ percent statement can be obtained from the right-hand inequality for determining, in the case of $q + r = p$, whether the last r roots are probably negligible. Anderson (1963a) has also given a more conservative confidence interval than (13), for that expression tends to be narrow when the population roots are in fact not equal.

Let us state an asymptotic test of the important hypothesis

$$H_0: \quad \boldsymbol{\alpha}_i = \boldsymbol{\alpha}_{i0} \tag{15}$$

that the characteristic vector associated with the *distinct* root λ_i of $\boldsymbol{\Sigma}$ is equal to some specified vector $\boldsymbol{\alpha}_{i0}$. As in the case of the preceding tests and interval estimates, the result is due to Anderson (1963a). Starting from the result (3) and its covariance matrix (6) it can be shown that the test statistic

$$\chi^2 = n\left(l_i\boldsymbol{\alpha}'_{i0}\mathbf{S}^{-1}\boldsymbol{\alpha}_{i0} + \frac{1}{l_i}\boldsymbol{\alpha}'_{i0}\mathbf{S}\boldsymbol{\alpha}_{i0} - 2\right) \tag{16}$$

is asymptotically distributed as a chi-squared variate with $p - 1$ degrees of freedom when H_0 is true.

Jolicoeur (1984) has proposed an exact small-sample test of the hypothesis (15) based on the multiple correlation of the ith population principal-component values with those of the other $p - 1$ population components. Since the component variates are uncorrelated, (15) would be rejected if the associated F statistic

$$F = (N - p)[(\boldsymbol{\alpha}'_{i0}\mathbf{S}\boldsymbol{\alpha}_{i0})(\boldsymbol{\alpha}'_{i0}\mathbf{S}^{-1}\boldsymbol{\alpha}_{i0}) - 1]/(p - 1) \tag{17}$$

exceeds the critical value $F_{\alpha;p-1,N-p}$. Jolicoeur's test does not attach the hypothesized vector $\boldsymbol{\alpha}_{i0}$ to any particular characteristic root of $\mathbf{S}$.

In many principal-component analyses the later characteristic roots have been observed to decline linearly, rather than essentially converge monotonically to some constant. Cattell (1966) remarked on that property for correlation matrices in his proposal of the scree plot. Bentler and Yuan (1996) have constructed a generalized likelihood-ratio test for the hypothesis of a linear trend in the last q characteristic roots of a covariance matrix. The test requires the numerical solution of certain nonlinear estimating equations. Bentler and Yuan have applied it to the characteristic roots of two covariance matrices of psychological test scores discussed repeatedly in the psychometric literature, and have confirmed the linear trends in the later component variances.

Example 6.11. *Allometry* is the study of the relationship of the size and shape of an organism, as well as the relative growth of its various dimensions. Its origins extend back to Galileo and his investigations of the limiting sizes of organisms and structures. The concept was introduced in modern biology by Huxley (1924, 1932). Gould (1966) has surveyed its biological role, while its statistical and mathematical properties have been described by a number of workers, including Jolicoeur (1963b), Mosimann (1970), and Sprent (1972). Niklas (1994) has discussed allometric properties of plants. West *et al.* (1997, 1999) have given a general fractal geometry model for allometric scaling laws based on the physics and hydrodynamics of capillaries and tubes. MacKenzie

(1999) has given an overview of the work of West and his collaborators, and descriptions of some competing allometric models. Jolicoeur has proposed this generalization of allometry to the growth relations of p dimensions of an organism: The growth of the dimensions is isometric, or in constant proportion with increasing size, if the first principal component of the covariance matrix of the logarithms of the dimensions has the equiangular direction-cosine vector

$$\alpha_1' = \left[\frac{1}{\sqrt{p}}, \ldots, \frac{1}{\sqrt{p}}\right]$$

Let us test the hypothesis of isometric growth model for the turtle carapace data introduced in Example 6.2. The covariance matrix of the logarithms of the three measurements for the 24 turtles and its first characteristic vector are given in Example 6.4. The inverse matrix is

$$\mathbf{S}^{-1} = \begin{bmatrix} 5.62396 & -3.54956 & -2.82829 \\ & 7.15698 & -2.00277 \\ & & 4.54773 \end{bmatrix}$$

and the statistic (16) has the value 31.71. Since this is much greater than the upper 1% critical value of the chi-squared distribution with 2 degrees of freedom, the isometric growth is untenable for the turtle carapaces.

Waternaux (1976) and Davis (1977) have studied the robustness of principal-component distributions to nonnormality in the original observations. Waternaux has expressed the large-sample normal distribution of distinct characteristic roots in terms of the fourth-order cumulants of the underlying nonmultinormal population, and has made simulation studies of the means, variance, and confidence interval coverage proportions for short-, long-, and very-long-tailed trivariate populations. She concluded that tests or confidence intervals based on the distributional results given at the beginning of this section could be seriously affected by nonnormality. Davis has found the effects of nonnormality on the hypothesis tests and confidence intervals for characteristic roots and vectors, and has given conditions for inferences to be conservative. Dudzinski *et al.* (1975) have measured the repeatability of principal-component coefficients through simulation of samples from multinormal and nonmultinormal populations, and have concluded that repeatability depends largely on the separation of the characteristic roots rather than normality.

Principal Components Extracted from Correlation Matrices

The asymptotic distribution theory of the characteristic roots and vectors of correlation matrices is exceedingly more complicated than the preceding results, and will not be presented for want of space. However, two special cases lead to tractable results, and their importance merits consideration.

Lawley (1963) has proposed a test of the equality of the last $p-1$ characteristic roots of the population correlation matrix. Such a hypothesis is precisely equivalent to that of the equality of all $[p(p-1)]/2$ correlations. Let us begin with our usual assumption of this section that a random sample of N p-element observation vectors has been drawn from some multinormal population

$N(\mu, \Sigma)$. If the ijth element of Σ is $\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$, the null hypothesis is

$$H_0: \quad \rho_{ij} = \rho \tag{18}$$

for all subscripts $i \neq j$. Let the sample correlation of the ith and jth responses be denoted by r_{ij}. Then, following Lawley's notation, define these quantities:

$$n = N - 1$$

$$\lambda = 1 - \rho$$

$$\mu = \frac{(p-1)^2(1-\lambda^2)}{p-(p-2)\lambda^2}$$

$$\bar{r}_k = \frac{1}{p-1}\sum_{\substack{i=1 \\ i \neq k}}^{p} r_{ik}$$

$$\bar{r} = \frac{2}{p(p-1)}\sum_{i<j}\sum r_{ij}$$

$\bar{r}_k$ is the average correlation of the kth response with the other variates, and $\bar{r}$ is the grand mean of the correlations. λ is of course the second characteristic root of the population correlation matrix when H_0 is true, and is estimated by $\hat{\lambda} = 1 - \bar{r}$. Lawley has shown that as n tends to infinity, the test statistic

$$\chi^2 = \frac{n}{\hat{\lambda}^2}\left[\sum_{i<j}\sum (r_{ij} - \bar{r})^2 - \hat{\mu}\sum_{k=1}^{p}(\bar{r}_k - \bar{r})^2\right] \tag{19}$$

for H_0 is distributed according to the chi-squared distribution with $1/2(p+1)(p-2)$ degrees of freedom.

Brien *et al.* (1984) have proposed an alternative statistic for testing the hypothesis of a common correlation that is based on the Fisher z-transformed correlations. The statistic has the same large-sample chi-squared distribution as (19), but can be decomposed into two independent chi-squared components measuring, respectively, differences among the average row z values and departures from an additive model for row and column effects.

Example 6.12. We shall apply Lawley's statistic to the four-variable submatrix of wing- and leg-bone dimension correlations from the correlation matrix of Example 6.6. The mean correlations of the four variates are 0.8977, 0.9010, 0.8920, and 0.8960, and the grand mean correlation is $\bar{r} = 0.8967$; $n = 275$, $\hat{\lambda} = 0.10333$, and $\hat{\mu} = 2.2379$. The total and between-variates sums of squares are

$$\sum_{i<j}\sum (r_{ij} - \bar{r})^2 = 0.003943$$

$$\sum_{k=1}^{4}(\bar{r}_k - \bar{r})^2 = 0.0000420$$

and the test statistic is

$$\chi^2 = \frac{275}{0.01067}[0.003943 - (2.2379)(0.0000420)]$$
$$= 99.20$$

The $\alpha = 0.001$ critical value of the chi-squared distribution with 5 degrees of freedom is only 20.52, and we conclude that the hypothesis of a single major axis among the four skeletal measurements should be rejected.

The second case is that of a population correlation matrix with the two different roots λ_1 and λ_2 of respective multiplicities q_1 and q_2. The unit diagonal elements of the matrix imply that $\lambda_1 q_1 + \lambda_2 q_2 = p$. Let the characteristic roots of the sample matrix $\mathbf{R}$ be $l_1 > \cdots > l_{q_1} > \cdots > l_p$, and let the estimates of λ_1 and λ_2 be the averages

$$\bar{l}_1 = \frac{1}{q_1}\sum_{i=1}^{q_1} l_i \qquad \bar{l}_2 = \frac{1}{q_2}\sum_{i=q_1+1}^{p} l_i \tag{20}$$

As in the case of the population characteristic roots, $\bar{l}_2 = (p - q_1\bar{l}_1)/q_2$, and it will suffice to consider only one estimate, say $\bar{l}_2$, in the sequel. Anderson has shown that the asymptotic expectation is λ_2, its asymptotic variance is

$$2\lambda_2^2 \frac{(p - q_2\lambda_2)^2}{npq_1q_2} \tag{21}$$

and that as n becomes large, the standardized quantity

$$\frac{\bar{l}_2 - \lambda_2}{\lambda_2(p - q_2\lambda_2)}\sqrt{\frac{npq_1q_2}{2}} \tag{22}$$

tends to be distributed as a unit normal variate. From these results we can construct confidence intervals for either population root. For the $100(1-\alpha)$ percent confidence interval for λ_2 we require that the square of the standardized estimate (22) be less than $z_{1/2\alpha}^2$, the square of the appropriate unit normal percentage point. This inequality leads to two quadratic equations in λ_2. The appropriate roots give the confidence interval

$$\frac{Kp + 1 - \sqrt{(Kp+1)^2 - 4Kq_2\bar{l}_2}}{2Kq_2} \le \lambda_2 \le \frac{Kp - 1 + \sqrt{(Kp-1)^2 + 4Kq_2\bar{l}_2}}{2Kq_2} \tag{23}$$

where we have introduced the substitution $K = z_{1/2\alpha}\sqrt{2/npq_1q_2}$ in the interests of simplicity. A similar interval can be obtained for the larger characteristic root, although once again the reader is cautioned that the roots and their estimates are linearly dependent. That interval is given by expression (23) with q_2 replaced by q_1 and $\bar{l}_2$ replaced by $\bar{l}_1$.

Example 6.13. In the preceding discussion if $q_1 = 1$ and $q_2 = p - 1$, we have the especially important case of the equicorrelation matrix. The matrix

$$\begin{bmatrix} 1 & 0.9740 & 0.9726 \\ & 1 & 0.9655 \\ & & 1 \end{bmatrix}$$

of the correlations of the carapace data of Example 6.2 certainly appears to have come from such a population, and it will be of interest to compute the 95% confidence interval for the double characteristic root. Here $p = 3$, $q_2 = 2$, $n = 23$, and $z_{1/2\alpha} = 1.96$, and it can be shown that the greatest root $l_1 = 2.9414$. Thus $\bar{l}_2 = 1/2(3 - 2.9414) = 0.0293$. $K = 0.236$, and the confidence interval is

$$0.017 \leq \lambda_2 \leq 0.088$$

If we had preferred, we might have obtained instead the 95% confidence interval for the largest root:

$$2.824 \leq \lambda_1 \leq 2.966$$

The true 95 percent confidence intervals for a small sample of 24 observations would be somewhat wider.

Schott (1991) has constructed a large-sample chi-squared test for the hypothesis that the ith characteristic vector of a correlation matrix has the specified form α_0. The matrix of the test statistic quadratic form is the generalized inverse of a complicated function of the sample characteristic roots and vectors, and will be omitted here.

6.7 Some Further Topics

Influence Functions for Principal Components

It is often of interest to know which sampling units or response variables have especially strong effects on the size of characteristic roots or the elements of the characteristic vectors. *Influence functions* have been developed by Radhakrishnan and Kshirsagar (1981) and Critchley (1985) for measuring those effects. Several alternative measures have been proposed, but we shall only consider some simple and convenient ones based on the individual sampling unit principal-component scores defined by expression (25), Section 6.2. The jth component score for the ith unit is

$$y_{ij} = \mathbf{a}'_j(\mathbf{x}_i - \bar{\mathbf{x}}) \tag{1}$$

The influence function for the jth characteristic root l_j evaluated for the ith observation vector is

$$I(\mathbf{x}_i, l_j) = y_{ij}^2 - l_j \tag{2}$$

Note that the influence is negative if the score is less than the standard deviation $\sqrt{l_j}$. Larger values of the scores give a positive influence measure.

The influence function for the jth characteristic vector is

$$I(\mathbf{x}_i, \mathbf{a}_j) = -y_{ij} \sum_{\substack{h=1 \\ h \neq j}}^{p} y_{ih}\mathbf{a}_h/(l_h - l_j) \tag{3}$$

That measure is a function of all characteristic vectors except $\mathbf{a}_j$, and vanishes when $y_{ij} = 0$. If some of the characteristic roots are nearly equal to l_j the influence measure will increase without bound. The number of influence measures for the complete set of components, Np^2, is unfortunately very large.

Example 6.14. Principal components were extracted from the covariance matrix of the first six course evaluation scales of the data set used in Example 6.1. These characteristic roots were obtained by MINITAB and the APL SYMEIG function:

Index	**1**	**2**	**3**	**4**	**5**	**6**
Root	2.9821	0.6859	0.4200	0.1768	0.1045	0.0182

A scree plot suggests that only the first principal component, which explains 68.0% of the total variance, should be interpreted. The influence measures for the first characteristic root are given in the table.

Student	**Influence $\mathbf{I}(\mathbf{x}_i, \mathbf{l}_j)$**
1	3.28
2	−2.55
3	2.56
4	−2.63
5	−2.94
6	2.56
7	−1.86
8	2.56
9	−0.53
10	−1.96
11	0.69
12	−2.33
13	−2.94
14	−1.71
15	7.77
16	−2.95

The large influences of Students 1, 3, 6, 8, and 15 are associated with extreme values of the first component. The influence functions for the component coefficients will be omitted.

Jackson (1991) has illustrated these influence functions, and has given influence functions for characteristic roots and vectors computed from correlation matrices. Those functions are markedly more complex. Their derivations are due to Calder (1986), Jolliffe (1986), and Pack *et al.* (1988).

Biplots

Because the visualization of multivariate data with more than two or three responses is difficult Gabriel (1971) has proposed the *biplot* as one concise way of representing sample observations. The method assumes that the data effectively lie mainly in a two-dimensional space, or equivalently, that the first two principal components explain most of the variation. The representation is based on the factorization

(4) $$\mathbf{X} = \mathbf{G}\mathbf{H}'$$

of the $N \times p$ matrix $\mathbf{X}$ of rank r into the product of the $N \times r$ matrix $\mathbf{G}$ and the $r \times p$ matrix $\mathbf{H}'$. The latter two matrices are of rank r (Rao, 1973). In a sense, this is a multiplicative model for $\mathbf{X}$, where $\mathbf{G}$ represents row, or sampling unit, effects, and the columns of $\mathbf{H}'$ give a basis for the response-variable effects. For the usual biplot, $r = 2$. The decomposition (4) is of course not unique, for postmultiplication of $\mathbf{G}$ by any nonsingular $r \times r$ matrix and premultiplcation of $\mathbf{H}'$ by its inverse will give another decomposition. The result is an example of the Eckart-Young singular decomposition theorem (Eckart and Young, 1939).

In practice $\mathbf{X}$ is replaced by the matrix $\mathbf{Z}$ of the deviations $z_{ij} = x_{ij} - \bar{x}_j$ of the observations from their column means. The least-squares estimate of $\mathbf{G}$ is the matrix

(5) $$\mathbf{Y} = \mathbf{Z}[\mathbf{a}_1, \ldots, \mathbf{a}_p]$$

of the values of the p components. The least-squares estimate of $\mathbf{H}$ is the matrix of characteristic vectors

(6) $$\mathbf{H} = [\mathbf{a}_1, \ldots, \mathbf{a}_p]$$

For the biplot only the first two principal components are used, so that the approximation to $\mathbf{Z}$ is

(7) $$\hat{\mathbf{Z}} = [\mathbf{y}_1\ \mathbf{y}_2]\begin{bmatrix} \mathbf{a}'_1 \\ \mathbf{a}'_2 \end{bmatrix}$$

For the biplot the pairs of component values $(y_{i1},\ y_{i2})$ are plotted as points characterizing the individual sampling units, and the pairs of component coefficients $(a_{j1},\ a_{j2})$ are plotted as arrows from the origin of the coordinate axes. Usually the scales of the y_{ij} and the a_{jh} are different, and either separate axis scales must be specified, or the coordinate values scaled to improved the readability of the plot.

Biplots appear to be most useful for data with two very dominant characteristic roots. Large numbers of response variables and a large sample size may lead to a cluttered plot that is difficult to interpret. The biplot graphics option is available in the JMP 4 statistical system. Unfortunately, its plot size is small and the graphic is displayed as a negative white-on-black image.

Example 6.15. A biplot was prepared for the first $p = 6$ variables in the course evaluation data of Example 6.1, and is shown in Figure 6.4. These are the response variables and their symbols in the biplot:

1. Overall evaluation of the instructor (IO)

2. Overall evaluation of the course (CO)

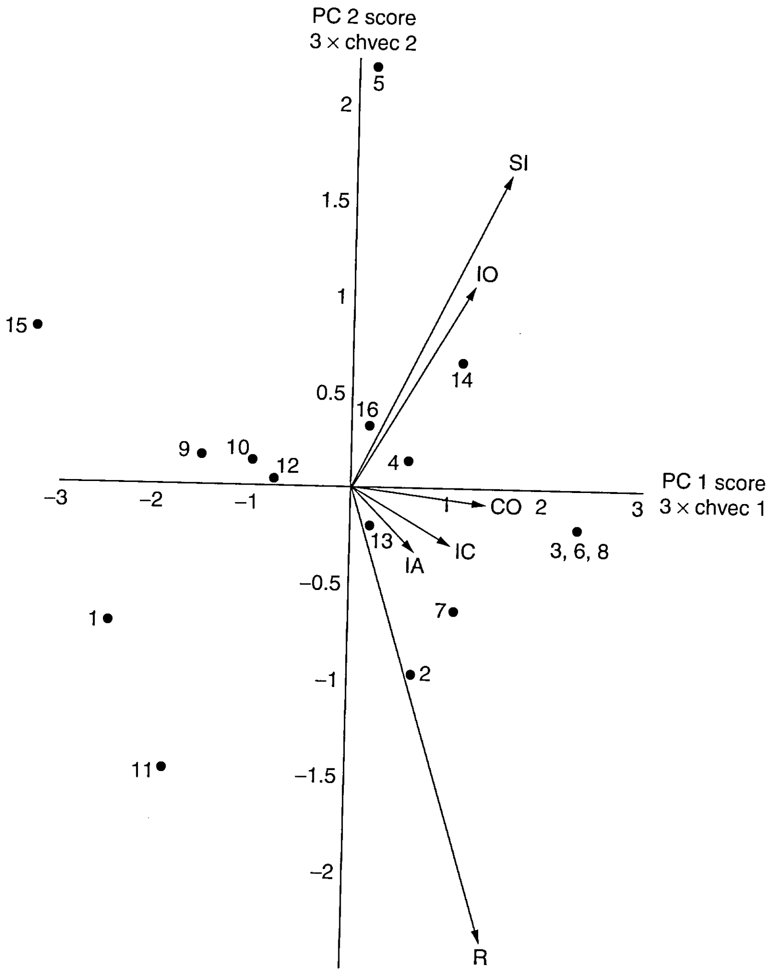

FIGURE 6.4 Biplot of the $p = 6$ course evaluation variables

3. Ability of the instructor to communicate the material (IC)
4. Ability of the instructor to stimulate interest in the material (SI)
5. Accessibility of the instructor (IA)
6. Value of assigned readings (R)

The first two characteristic roots accounted for 68.0% and 15.6% of the total variance. The characteristic vectors with rounded elements follow:

$$\mathbf{a}_1' = [0.40, 0.44, 0.32, 0.53, 0.20, 0.47]$$

$$\mathbf{a}_2' = [0.34, -0.02, -0.09, 0.53, -0.10, -0.77]$$

The coefficients were multiplied by three for greater clarity in the plot.

In the plot the variables instructor overall (IO) and stimulation of interest (SI) are closely associated, and reflect the instructor's presentation ability. Value of readings (R) is nearly orthogonal to IO and SI, but closest to instructor accessibility (IA). The scales overall course (CO) and instructor communication ability (IC) have short vectors located in the midst of the others.

The component scores of the $N = 16$ respondents to the questionnaire are indicated by the numbered points on the plot. Student 5 may be an outlier on the first component scale, Student 1 on the second, and Student 11 on both.

Common Principal Components

Flury (1984, 1988) generalized principal-components analysis to the case of k populations with different covariance matrices $\Sigma_1, \ldots, \Sigma_k$, different sets of characteristic roots, and a common characteristic vector matrix $\mathbf{P}$. Then

$$\mathbf{P}'\Sigma_i\mathbf{P} = \mathbf{D}_i$$

where $\mathbf{D}_i$ is the diagonal matrix of characteristic roots of the ith covariance matrix. Flury found the maximum-likelihood estimates of $\mathbf{P}$ and the $\mathbf{D}_i$, and a generalized likelihood-ratio test of the hypothesis of a common orthonormal component matrix $\mathbf{P}$. The test is equivalent to that of the simultaneous diagonalizability of the k covariance matrices. The advantages of the common component model are the greater stability of the coefficients because of the larger sample size, and the smaller number of parameters in the single matrix $\mathbf{P}$ as opposed to k orthonormal matrices.

Flury (1984) applied his method to the Fisher iris data and the Jolicoeur North American marten bone dimensions. The hypothesis of a common component matrix was rejected for the iris measurements, and was tenable for male and female martens.

Gu and Fung (2001) have given influence functions for the common principal-components model.

The Ubiquity of Principal-Components Analysis

More than one hundred years after their conception by Karl Pearson and over sixty-five since their explication by Harold Hotelling, principal components continue to find new applications in statistical analysis and modeling. Although their original motivation was from psychology, they have found uses in many disciplines from physics to the neurosciences. This breadth is evident in the very comprehensive treatment by Jackson (1991), who has placed particular emphasis on engineering and industrial applications.

Principal components have been used to reduce large gene microarray databases to more manageable sizes. Liu *et al.* (2002) used *block principal-components analysis* to visualize and cluster genetic data from the National Cancer Institute 60 cell lines data set (NCI60). For block principal-components analysis the variables are first collected into similar groups, and the first few components extracted for each group. The covariance matrix of those component scores is then computed, and its principal components found. The final component variables are used to describe the original high-dimensional data set. Owens and King (1999) explained 75% of European DNA genetic variation by the first five principal components of allele frequencies, and were able to trace human migration patterns throughout Europe, Africa, and the Middle East.

Finlay and Darlington (1995) have proposed a two-component model for the natural logarithms of eleven brain structure models of 131 mammalian species. The first component accounted for 96.29% of the total variance, and is essentially brain size. The second component, after an orthogonal transformation, appeared to be related to the size of the limbic region and the olfactory bulb of the brain.

Doty *et al.* (1989) have used principal components to relate olfactory dysfunction in Parkinson's disease patients to cognitive and neurological test scores. The olfactory component appeared to be distinct from the components reflecting the other measures. Doty *et al.* (1994) concluded from the first four components of a data set of thirteen scores of odor perception tests administered to $N = 97$ human subjects that the scores measured a large portion of common variance, in addition to later dimensions identified with measures of intensity and pleasantness.

Dawkins (1989) used principal components to rank countries participating in the 1984 Olympic Games by the track records of men and women athletes. Because the variances of the times differed according to the length of the course Dawkins standardized the data and computed the components from the correlation matrix. The data of the men and women were analyzed separately, and biplots were used to display the components. Naik and Khattree (1996) reconsidered the data set by converting event times to speed, defined as "distance covered per unit time." Now each event has more commensurable velocities. The principal components of the new male and female data sets were used to rank the competing countries, with slightly different orderings than those obtained by Dawkins from the standardized data.

6.8 Exercises

1. Compare the first principal-component coefficients of the within-groups covariance matrix of Example 6.7 with the linear discriminant function coefficients of Example 4.1. To what extent could the first component substitute for the discriminant function?

2. In a psychological experiment[1] the reaction times of 64 normal men and women to visual stimuli were recorded when warning intervals of 0.5, 1, 3, 6, and 15 sec preceded the stimulus. The correlations of the median reaction times of several replications of each preparatory interval for a subject formed this matrix

$$\begin{bmatrix} 1 & 0.71 & 0.58 & 0.56 & 0.65 \\ & 1 & 0.71 & 0.60 & 0.69 \\ & & 1 & 0.75 & 0.71 \\ & & & 1 & 0.74 \\ & & & & 1 \end{bmatrix}$$

   (*a*) What is the greatest amount of variance that a single component could possibly explain?

   (*b*) Choose a trial vector consistent with the matrix pattern, and extract the first principal component.

[1] These correlations have been presented with the kind permission of Dr. Jack Botwinick.

(*c*) Compute the matrix reproduced by the first component, and the residual matrix.

(*d*) From the pattern of the residual matrix select a trial vector orthogonal to the coefficients of the first principal component, and extract the second characteristic vector.

(*e*) Test the hypothesis of equal correlations in the population out of which the subjects were drawn, and if it is tenable, construct the 99% confidence interval for the multiple smaller characteristic root.

3. In a study of the relations among the optical elements of the eye van Alphen (1961) computed the correlations of ocular refraction, axial length, anterior chamber depth, corneal power, and lens power of the right eyes of 886 young men and women whose axial lengths did not exceed ± 3 diopters. The five elements had this matrix of correlations:

$$\begin{bmatrix} 1 & -0.45 & -0.40 & -0.21 & 0.13 \\ & 1 & 0.45 & -0.52 & -0.60 \\ & & 1 & 0.09 & -0.32 \\ & & & 1 & -0.09 \\ & & & & 1 \end{bmatrix}$$

(*a*) Can the scales of the variates be reflected so that all correlations are positive?

(*b*) Extract the first principal component. Does the conjecture of a single dimension among the five elements appear to be justified?

(*c*) Compute the residual matrix of the first component. On the basis of its elements and the amount of variance explained by the first component would it appear that the second component can be computed easily?

4. Burton (1963) reexamined the data of Hartshorne and May (1928) in an investigation of the unidimensionality or generality of moral behavior. The correlations[2] of inventories measuring dishonesty in situations involving copying, speed, peeping, faking, athletic events, and lying had these values for one set of school children:

$$\begin{bmatrix} 1 & 0.450 & 0.400 & 0.400 & 0.288 & 0.350 \\ & 1 & 0.374 & 0.425 & 0.345 & 0.248 \\ & & 1 & 0.300 & 0.100 & 0.108 \\ & & & 1 & 0.300 & 0.256 \\ & & & & 1 & 0.230 \\ & & & & & 1 \end{bmatrix}$$

(*a*) Verify that the correlation coefficients of the measures with the first two components are given by the vectors

$$\mathbf{a}_1' = [0.764, 0.754, 0.581, 0.703, 0.555, 0.526]$$

$$\mathbf{a}_2' = [0.092, 0.106, 0.660, 0.017, -0.504, -0.504]$$

(*b*) What percentage of the total standard scores variance do these components explain?

5. Use the representation

$$\Sigma = \sigma^2[1 + (p-1)\rho]\alpha_1\alpha_1' + \sigma^2(1-\rho)\sum_{h=2}^{p}\alpha_h\alpha_h'$$

of the equal-variance, equal-correlation matrix to verify the variance and covariance of expression (8), Section 6.6.

6. In an allometric investigation of the North American marten Jolicoeur (1963a) measured the following dimensions of the humerus and femur bones of 92 males and 47 females:

X_1 = humerus length from head to medial condyle
X_2 = maximum epicondylar width of the distal end of the humerus
X_3 = femur length from head to medial condyle
X_4 = maximum width of the femur distal end

The diagonal matrix of the characteristic roots of the within-sex covariance matrix $\mathbf{S}$ of the logarithms of those observations was

$$\hat{\Lambda} = 10^{-4}\begin{bmatrix} 4.2813 & 0 & 0 & 0 \\ & 1.2603 & 0 & 0 \\ & & 0.5738 & 0 \\ & & & 0.1272 \end{bmatrix}$$

and the matrix formed from the column vectors of direction cosines was

$$\mathbf{A} = \begin{bmatrix} 0.39410 & 0.50972 & 0.22651 & 0.73046 \\ 0.57374 & -0.52333 & 0.61536 & -0.13518 \\ 0.39568 & 0.63427 & 0.02469 & -0.66373 \\ 0.59913 & -0.25301 & -0.75459 & 0.08731 \end{bmatrix}$$

Use the relationships $\mathbf{S} = \mathbf{A}\hat{\Lambda}\mathbf{A}'$ and $\mathbf{S}^{-1} = \mathbf{A}\hat{\Lambda}^{-1}\mathbf{A}'$ to test the hypothesis

$$H_0: \quad \alpha_1' = \left[\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right]$$

of equal direction cosines, or the same relative growth rates, for the dimensions in the *Martes americana* population.

7. Weiss (1963) reported this matrix of correlations among the hearing threshold measures left-ear audiogram hearing loss, right-ear audiogram hearing loss, 2,000 clicks/sec threshold, and 1 click/sec threshold:

$$\begin{bmatrix} 1 & 0.81 & -0.82 & -0.80 \\ & 1 & -0.91 & -0.80 \\ & & 1 & 0.86 \\ & & & 1 \end{bmatrix}$$

The sample consisted of $N = 43$ elderly men.

(*a*) Extract the first principal component from this matrix.

(b) Change the signs of the last two responses so that all correlations are positive. Carry out Lawley's test for the hypothesis of equality of the six population correlations among the four measures.

(c) Construct the 95 percent confidence interval for the greatest population characteristic root.

(d) If the hypothesis tested in part (b) can be accepted, construct the 95 percent confidence interval for the smaller characteristic root of the matrix of positive correlations.

8. The correlation matrix of the course evaluation data described in Exercise 17, Chapter 1, is

$$\mathbf{R} = \begin{bmatrix} 1 & 0.88 & 0.95 & 0.80 \\ & 1 & 0.85 & 0.71 \\ & & 1 & 0.82 \\ & & & 1 \end{bmatrix}$$

(a) How could you use the results of Section 6.5 to infer the approximate nature of the principal components of **R**?

(b) These principal components were extracted from **R**. What interpretations might be given to the components?

Variable	Component 1	2	3	4
1. Overall evaluation of instructor	0.52	0.18	0.36	0.75
2. Ability to present course material	0.49	0.54	−0.66	−0.15
3. Instructor's ability to stimulate interest	0.52	0.05	0.56	−0.64
4. Overall evaluation of course	0.47	−0.82	−0.33	0.03
Characteristic root	3.51	0.30	0.15	0.04

9. Another course evaluation data set consisted of responses from $N = 32$ M.B.A. students in a basic statistics class. The correlation matrix of six of the scales had these components:

Variable	Component 1	2	3	4	5	6
1. Presentation	0.42	0.43	0.49	0.07	0.25	−0.57
2. Ability to stimulate interest	0.41	0.37	−0.66	−0.25	−0.41	−0.20
3. Knowledge of material	0.27	−0.55	0.31	−0.67	−0.24	−0.13
4. Interest in students	0.40	−0.45	0.43	0.12	0.65	−0.11
5. Objectivity	0.41	−0.33	0.13	0.67	−0.51	0.02
6. Overall evaluation of instructor	0.51	0.26	0.18	−0.12	0.15	0.78
Characteristic root	3.20	1.50	0.55	0.45	0.22	0.08

(a) Interpret as many components as appear to explain appreciable proportions of the total variance.

(b) Which components could be omitted, based on the sizes or lack of uniqueness of their variances?

(c) Reproduce the correlation matrix from the characteristic roots and vectors. Compare the correlations with those reproduced using only the first two components and their characteristic roots.

10. Does the equipredictability matrix (8) of Section 6.5 have the Type H pattern described in Section 3.6 for the profile analysis of variance?

11. The correlation matrix

$$\mathbf{P} = \begin{bmatrix} 1 & \rho & \rho/2 & \rho/4 \\ & 1 & \rho & \rho/2 \\ & & 1 & \rho \\ & & & 1 \end{bmatrix}$$

arises in quantitative genetics as the intergenerational correlations of a trait along a direct line of descent, *e.g.*, from great-grandmother to daughter (Rutledge, 1976).

(a) Show that the characteristic roots and vectors of $\mathbf{P}$ are the following:

Root	Vector
$1 + 2.1712\rho$	[0.4352, 0.5573, 0.5573, 0.4352]
1	[−2, −1, 1, 2]
$1 - 0.9212\rho$	[−0.5573, 0.4352, 0.4352, −0.5573]
$1 - \frac{5}{4}\rho$	[1, −2, 2, −1]

(b) What are the limits of ρ if $\mathbf{P}$ is to be positive semidefinite?

12. Test the hypotheses

$$H_0: \quad \mathbf{a}_1' = \left[\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right]$$

that the first principal axes of the no-senile-factor, senile-factor, and within-groups covariance matrices of Example 6.7 are the equiangular lines.

13. Jolicoeur (1963a, 1963c, 1984) investigated the skeletal dimensions of $N = 68$ male adult martens. The common logarithms of the skull, humerus, and femur lengths gave the following covariance matrix ($\times 10^4$) and principal components[3]:

$$10^4 \times \mathbf{S} = \begin{bmatrix} 0.996248 & 0.761328 & 0.752440 \\ & 1.311073 & 1.168278 \\ & & 1.319772 \end{bmatrix}$$

Variable	Component 1	Component 2	Component 3
Skull length	0.463356	0.885872	−0.023049
Humerus length	0.626388	−0.309013	0.715646
Femur length	0.626848	−0.346036	−0.698083
Characteristic root ($\times 10^4$)	3.043383	0.436763	0.146946

(a) Test the isometry hypothesis

$$H_0: \quad \alpha_1' = [\sqrt{3}/3, \sqrt{3}/3, \sqrt{3}/3]$$

by the Anderson and Jolicoeur statistics of Section 6.6.

[3] Reproduced from: P. Jolicoeur, "Principal Components, Factor Analysis, and Multivariate Allometry: A Small-Sample Direction Test", *Biometrics*, vol. 40, pp. 685–690, 1984. With permission of the Biometric Society.

(*b*) Jolicoeur has pointed out that natural selection would have led to minimum variation in the direction of the vector

$$\alpha_3' = [0, \sqrt{2}/2, -\sqrt{2}/2]$$

defining the contrast of the logarithms of forelimb and hindlimb dimensions. Test the hypothesis that the third population principal component has the coefficients of α_3.

14. Test the hypotheses of equal second, third, and fourth characteristic roots for the no-senile-factor, senile-factor, and within-groups covariance matrices of Example 6.7.

15. When principal components have been extracted from two independent sample covariance matrices with respective degrees of freedom $n_1 = N_1 - 1$ and $n_2 = N_2 - 1$, we may make an approximate test of

$$H_0: \quad \lambda_{i1} = \lambda_{i2}$$

of equal ith distinct population characteristic roots by referring

$$z = \frac{l_{i1} - l_{i2}}{\sqrt{\dfrac{2l_{i1}^2}{n_1} + \dfrac{2l_{i2}^2}{n_2}}}$$

to critical values of the standard normal distribution. Use the statistic to test the hypothesis of equal largest characteristic roots for the no-senile-factor and senile-factor component analyses of Example 6.7.

16. Closing common stock prices of six petroleum companies were collected for $N = 48$ trading days from May 4 to July 10, 1987. Summary statistics and the correlation matrix of the prices are shown in the displays (SD denotes standard deviation):

Company	**Mean**	**SD**	**Minimum**	**Maximum**
Exxon (XON)	89.51	2.51	86.625	94.875
Atlantic Richfield (ARC)	91.74	3.24	85.000	97.875
Chevron (CHV)	59.16	1.91	55.25	63.375
Amoco (AN)	86.14	1.85	82.125	89.875
Mobil (MOB)	49.31	1.90	46.00	52.625
Phillps (P)	16.23	0.69	14.75	17.875

$$\mathbf{R} = \begin{bmatrix} 1 & 0.8787 & 0.8866 & 0.7305 & 0.8303 & 0.7313 \\ & 1 & 0.8769 & 0.8863 & 0.9281 & 0.8404 \\ & & 1 & 0.7918 & 0.8533 & 0.8279 \\ & & & 1 & 0.8996 & 0.7727 \\ & & & & 1 & 0.7605 \\ & & & & & 1 \end{bmatrix}$$

Because much of the intercorrelation is due to a common upward trend in stock prices, linear and quadratic orthogonal polynomials were fitted to the data, and the residuals of each stock price from its predicted value were correlated. The trading days were treated as contiguous, so that holidays and weekends were ignored in fitting the polynomials. The standard deviation of the residuals and the squared multiple correlation coefficients of the stock prices with the two polynomials had these values:

	XON	ARC	CHV	AN	MOB	P
SD	0.9610	1.8490	0.8545	1.2877	1.2277	0.4402
R^2	0.85	0.67	0.80	0.52	0.58	0.59

The correlation matrix of the residuals is

$$\mathbf{R} = \begin{bmatrix} 1 & 0.5536 & 0.3528 & 0.3363 & 0.5071 & 0.0835 \\ & 1 & 0.5588 & 0.7815 & 0.8186 & 0.5758 \\ & & 1 & 0.5375 & 0.5915 & 0.4910 \\ & & & 1 & 0.8203 & 0.5445 \\ & & & & 1 & 0.4204 \\ & & & & & 1 \end{bmatrix}$$

(a) Test the null hypothesis that the six stock prices have the equal-correlation matrix. You must make the rather unrealistic assumption of independent daily prices for each stock.

(b) The principal components extracted from the correlation matrix of stock prices had these direction cosines and characteristic roots:

	Component					
Company	**1**	**2**	**3**	**4**	**5**	**6**
XON	0.40	−0.62	−0.31	−0.33	0.38	0.34
ARC	0.43	0.07	−0.09	−0.45	−0.12	−0.76
CHV	0.41	−0.38	0.08	0.78	−0.08	−0.26
AN	0.40	0.64	−0.14	0.20	0.60	0.08
MOB	0.42	0.25	−0.36	0.00	−0.69	0.40
P	0.39	0.03	0.86	−0.20	−0.06	0.25
Root	5.1699	0.3134	0.2839	0.1075	0.0784	0.0469

Interpret the first component. To what extent can the remaining five components be interpreted and given meanings?

(c) These components were extracted from the covariance matrix of the six stock prices:

	Component					
Company	**1**	**2**	**3**	**4**	**5**	**6**
XON	0.47	0.72	−0.15	0.43	0.22	−0.10
ARC	0.64	−0.23	−0.55	0.45	0.00	0.18
CHV	0.35	0.26	0.74	−0.44	−0.11	0.24
AN	0.33	−0.54	0.32	0.38	0.58	0.07
MOB	0.36	−0.26	0.12	0.42	−0.76	−0.20
P	0.12	−0.04	0.10	−0.31	0.15	−0.93
Root	25.0975	1.4557	0.5706	0.4335	0.2977	0.0959

Compare the first few components with those from the correlation matrix.

(d) The correlation matrix of the residuals from the fitted second-degree polynomial model had these components:

Company	1	2	3	4	5	6
			Component			
XON	0.30	0.75	−0.24	−0.44	0.30	−0.08
ARC	0.48	0.07	0.21	−0.20	−0.62	0.54
CHV	0.38	−0.12	−0.78	0.45	0.00	0.13
AN	0.45	−0.12	0.43	0.26	0.66	0.30
MOB	0.47	0.13	0.27	0.32	−0.27	−0.72
P	0.34	−0.62	−0.14	−0.62	0.11	−0.27
Root	3.7492	0.9553	0.5488	0.4507	0.1640	0.1320

Interpret the structure of those components that appear to be uniquely defined. To what extent has the removal of the polynomial trend changed the components of the stock prices?

17. The characteristic roots and vectors of the sums of squares and products matrices of the three iris species measurements given in Appendix B are shown in the table.

1. Iris setosa

Dimension	Characteristic vector 1	2	3	4
Sepal length	0.6691	0.5979	−0.4400	0.0361
Sepal width	0.7341	−0.6207	0.2746	0.0196
Petal length	0.0965	0.4901	0.8324	0.2399
Petal width	0.0636	0.1309	0.1951	−0.9699
Characteristic root	11.5863	1.8090	1.3130	0.4426

2. Iris versicolor

Dimension	Characteristic vector 1	2	3	4
Sepal length	0.6867	−0.6691	0.2651	0.1023
Sepal width	0.3053	0.5675	0.7296	−0.2289
Petal length	0.6237	0.3433	−0.6272	−0.3160
Petal width	0.2150	0.3353	−0.0637	0.9150
Characteristic root	23.9058	3.5468	2.6840	0.4797

3. Iris virginica

Dimension	Characteristic vector 1	2	3	4
Sepal length	0.7410	−0.1653	0.5345	0.3714
Sepal width	0.2033	0.7486	0.3254	−0.5407
Petal length	0.6279	−0.1694	−0.6515	−0.3906
Petal width	0.1238	0.6193	−0.4290	0.6459
Characteristic root	34.0675	5.2210	2.5625	1.6790

(*a*) Compare the three principal-component structures of the iris species. Do the third and fourth components appear to be negligible in their explanation of the total variances?

(*b*) Compare some of the characteristic roots of the three species. What hypothesis tests might be carried out on pairs of roots from the three component analyses?

(c) Compute the values of the first two components for each of the flowers in the three species samples. Can the component scores or their plots be used to discriminate among the three species?

(d) Extract the characteristic roots and vectors from the correlation matrices of the iris dimensions for the three species. Compare the component structures with those given in the table for the sums of squares and products matrices.

18. Principal components were extracted from the covariance and correlation matrices of the $p = 19$ aphid dimensions given in Appendix B. The first two components and their variances are given in the table. The next six characteristic roots and the traces of each matrix are shown below for information; their characteristic vectors have been omitted:

	Characteristic root						
	3	4	5	6	7	8	Trace
Covariance matrix	1.02	0.74	0.17	0.12	0.11	0.08	30.0594
Correlation matrix	0.75	0.50	0.28	0.26	0.18	0.16	19

	Matrix			
	Covariance		Correlation	
Variable	1	2	1	2
1	0.77	0.09	0.25	0.01
2	0.39	0.11	0.26	0.07
3	0.25	0.03	0.26	0.03
4	0.17	0.06	0.26	0.09
5	0.04	−0.10	0.16	−0.40
6	0.06	−0.04	0.24	−0.18
7	0.06	−0.04	0.25	−0.16
8	0.08	0.08	0.23	0.24
9	0.07	0.01	0.24	0.04
10	0.07	−0.01	0.25	−0.03
11	−0.12	−0.16	−0.13	−0.21
12	0.12	0.00	0.26	0.01
13	0.15	0.02	0.26	0.03
14	0.12	0.03	0.26	0.07
15	0.17	0.00	0.25	−0.01
16	0.09	−0.17	0.20	−0.40
17	0.16	−0.94	0.11	−0.55
18	−0.06	−0.12	−0.19	−0.35
19	0.07	0.10	0.20	0.28
Characteristic root	23.9908	3.5692	13.8383	2.3678

(a) Jeffers (1967) concluded that the effective dimensionality of the aphid data was 2. How do the preceding characteristic roots support that conclusion?

(b) Compare the corresponding characteristic vectors of the covariance and correlation matrices. How has standardization affected the component structure?

(c) Prepare scree plots for the eight characteristic roots of each matrix and use them to determine the number of components that should be interpreted in each analysis.

19. Jolicoeur (1963a) gave the covariance matrix of the logarithms of four bone dimensions of $N = 92$ North American male martens. The bone dimensions were humerus length (HL), femur length (FL), humerus width (HW), and femur width (FW). This covariance matrix (multiplied by 10^4) and correlation matrix were obtained:

$$10^4 \times \mathrm{S} = \begin{bmatrix} 1.1544 & 1.0330 & 0.9109 & 0.7993 \\ & 1.2100 & 0.7056 & 0.7953 \\ & & 2.0381 & 1.4083 \\ & & & 2.0277 \end{bmatrix}$$

$$\mathrm{R} = \begin{bmatrix} 1 & 0.8740 & 0.5939 & 0.5224 \\ & 1 & 0.4493 & 0.5077 \\ & & 1 & 0.6928 \\ & & & 1 \end{bmatrix}$$

The covariance matrix had these principal components expressed in direction cosine form:

	Component			
Dimension	**1**	**2**	**3**	**4**
HL	0.412	0.521	0.148	0.733
FL	0.389	0.641	−0.128	−0.649
HW	0.585	−0.402	0.681	−0.181
FW	0.580	−0.395	−0.706	0.097
Variance	4.548	1.116	0.645	0.121

(a) Test the hypothesis of equal correlations.

(b) Test the hypothesis of an equiangular first principal component.

20. MINITAB Release 12 was used to extract the principal components from the covariance and correlation matrices of the boy and girl achievement test scores of Exercise 38, Chapter 2. To compute the component coefficients and variances one clicks on the Multivariate statistical methods, and then on the Principal Components option. The variable columns are then specified, and the correlation or covariance matrix source is chosen. These sets of six complete components were found for boys and girls and the two matrices:

Girls:

Covariance matrix

	Component					
Test score	**1**	**2**	**3**	**4**	**5**	**6**
V	0.524	−0.446	0.006	−0.473	0.522	0.174
WREC	0.387	−0.029	0.866	0.213	−0.182	−0.146
RC	0.474	−0.304	−0.365	0.475	−0.426	0.378
MCON	0.360	0.251	−0.183	−0.586	−0.567	−0.329
MCOMP	0.316	0.791	0.004	0.039	0.240	0.464
MPS	0.349	0.139	−0.290	0.400	0.366	−0.694
Variance	8591.0	1231.4	719.4	515.7	348.7	248.1
Proportion	0.737	0.106	0.062	0.044	0.030	0.021

Correlation matrix

Test score	Component 1	2	3	4	5	6
V	0.422	−0.425	−0.039	0.233	0.683	−0.346
WREC	0.395	−0.183	0.847	−0.139	−0.147	0.229
RC	0.424	−0.360	−0.309	−0.166	−0.644	−0.390
MCON	0.417	0.217	−0.162	0.754	−0.196	0.383
MCOMP	0.361	0.777	0.119	−0.137	0.072	−0.477
MPS	0.426	0.078	−0.382	−0.559	0.232	0.549
Variance	4.3632	0.6375	0.3903	0.3061	0.1554	0.1476
Proportion	0.727	0.106	0.065	0.051	0.026	0.025

Boys:

Covariance matrix

Test score	Component 1	2	3	4	5	6
V	0.509	0.099	−0.637	0.298	0.464	0.147
WREC	0.395	0.074	0.124	0.354	−0.252	−0.797
RC	0.467	0.197	−0.115	−0.836	−0.145	−0.094
MCON	0.575	−0.553	0.400	0.120	−0.168	0.402
MCOMP	0.093	0.150	0.578	−0.110	0.776	−0.144
MPS	0.169	0.786	0.268	0.245	−0.265	0.390
Variance	4040.4	1492.6	1273.2	742.1	462.2	360.7
Proportion	0.483	0.178	0.152	0.089	0.055	0.043

Correlation matrix

Test score	Component 1	2	3	4	5	6
V	0.449	−0.375	−0.295	0.010	0.741	0.149
WREC	0.517	0.000	0.033	−0.554	−0.166	−0.631
RC	0.474	−0.085	−0.129	0.767	−0.352	−0.200
MCON	0.451	−0.162	0.562	−0.160	−0.251	0.605
MCOMP	0.188	0.710	0.448	0.211	0.446	−0.125
MPS	0.261	0.567	−0.616	−0.187	−0.194	0.398
Variance	2.6463	1.2262	0.9141	0.5514	0.3793	0.2827
Proportion	0.441	0.204	0.152	0.092	0.063	0.047

(a) Verify some of the principal-component coefficients and variances by computing them from the original observations by MINITAB or another statistical software system.

(b) Does the second principal component in the female data set have a fairly straightforward interpretation?

(c) In the female scores is most of the variation explained by the first few components, so that the last components might be ignored?

(d) What gender differences seem to be apparent through the principal-component structures?

21. Several principal-component analyses were carried out on the cerebral glucose metabolic rate data of Exercise 26, Chapter 1. The characteristic roots and vectors were extracted by the MINITAB Release 12 multivariate principal-components subroutine. In each case the coefficients of the first components had negative signs; these were dropped for simplicity. The results follow:

Complete covariance matrix:

	Component				
Variable	**1**	**2**	**3**	**4**	**5**
WBDIR	0.427	−0.083	0.232	−0.027	0.870
LCALCA	0.507	0.568	−0.108	0.623	−0.146
RCS	0.413	−0.426	0.682	0.072	−0.422
LMP	0.469	−0.547	−0.685	−0.025	−0.100
RSP	0.414	0.435	−0.009	−0.778	−0.183
Eigenvalue	11.118	0.172	0.142	0.063	0.005
Variance proportion	0.967	0.015	0.012	0.005	0.000

Complete correlation matrix:

	Component				
Variable	**1**	**2**	**3**	**4**	**5**
WBDIR	0.454	0.135	0.155	−0.007	0.867
LCALCA	0.448	−0.475	0.073	−0.733	−0.179
RCS	0.445	0.653	0.452	−0.005	−0.415
LMP	0.444	0.221	−0.858	0.070	−0.112
RSP	0.447	−0.530	0.173	0.677	−0.177
Eigenvalue	4.8357	0.0745	0.0608	0.0265	0.0026
Variance proportion	0.967	0.015	0.012	0.005	0.001

Four-variable covariance matrix:

	Component			
Variable	**1**	**2**	**3**	**4**
LCALCA	0.561	0.540	−0.102	0.619
RCS	0.455	−0.382	0.803	0.053
LMP	0.519	−0.621	−0.587	−0.025
RSP	0.458	0.421	−0.007	−0.783
Eigenvalue	9.0973	0.1706	0.1324	0.0630
Variance proportion	0.961	0.018	0.014	0.007

Four-variable correlation matrix:

	Component			
Variable	**1**	**2**	**3**	**4**
LCALCA	0.503	0.454	0.053	0.734
RCS	0.497	−0.635	0.591	0.010
LMP	0.498	−0.347	−0.791	−0.070
RSP	0.502	0.520	0.148	−0.675
Eigenvalue	3.8417	0.0730	0.0588	0.0265
Variance proportion	0.960	0.018	0.015	0.007

(a) Verify one set of principal components using MINITAB or another statistical program.

(b) Does the similarity of the covariance and correlation components reflect the relative magnitudes of the original variances? Why or why not?

(c) The last four variables were used as predictors of WBDIR in a multiple regression model. Do their eigenvalues suggest that they are nearly collinear?

(d) Test the hypotheses that the first characteristic vectors of the population covariance matrices are the equiangular vectors for $p = 5$ and $p = 4$, respectively.

(e) Use the correlation matrix of the original Exercise 26, Chapter 1, to test the hypothesis of an equal-correlation matrix in the population.

22. Anderson (1963a) has proposed the covariance matrix model $\boldsymbol{\Sigma} = \boldsymbol{\Psi} + \sigma^2\mathbf{I}$ for multiple measurements of a quantity, where the true measurements have the covariance matrix $\boldsymbol{\Psi}$, and the measurement errors are independently distributed of the true measurements with variance σ^2. The case of $\boldsymbol{\Psi} = \mathbf{aa}'$, where $\mathbf{a}$ is a $p \times 1$ non-null vector, is especially important.

(a) Show that the first characteristic root of $\boldsymbol{\Sigma}$ is $\lambda = \sigma^2 + \mathbf{a}'\mathbf{a}$.

(b) What are the remaining $p - 1$ characteristic roots and vectors?

23. The four variates $X_1, \ldots, X_4$ have the covariance matrix

$$\begin{bmatrix} 8 & 4 & 2 & 1 \\ & 8 & 1 & 2 \\ & & 8 & 4 \\ & & & 8 \end{bmatrix}$$

(a) Verify that $Y_1 = X_1 + \cdots + X_4$ is the first principal component of the four-variate complex.

(b) Find the variance of Y_1.

(c) What is the correlation of X_4 and Y_1?

(d) What are the angles of Y_1 with the four original coordinate axes?

24. The wolf skull measurements of Exercise 5, Chapter 4, were reanalyzed by extracting principal components from the within-groups covariance matrix. The first three components had these direction cosines and characteristic roots:

Skull dimension	Component 1	2	3
1. Palatal length	0.187	0.317	0.427
2. Postpalatal length	0.258	0.615	−0.228
3. Zygomatic width	0.790	−0.150	−0.391
4. Palatal width outside first upper molars	0.187	0.073	0.444
5. Palatal width inside second upper molars	0.148	0.151	0.343
6. Width between postglenoid foramina	0.286	0.177	0.398
7. Interorbital width	0.331	−0.264	−0.060
8. Least width of brain case	0.162	−0.608	0.369
9. Crown length of first upper molar	0.012	−0.005	0.024
Characteristic root	39.8838	9.7218	5.2253

Sum of all characteristic roots: 65.7777

These component scores were computed for the individual wolves:

Rocky Mountain

	Males			Females		
	PC1	PC2	PC3	PC1	PC2	PC3
	225	64.7	61.6	211	65.8	57.8
	237	69.2	58.6	210	66.3	58.9
	236	66.4	61.5	205	64.7	60.2
	226	69.2	62.7			
	227	68.2	60.9			
	226	72.3	59.1			
Mean	229.40	68.33	60.73	208.43	65.63	58.39

Arctic

	Males			Females		
	PC1	PC2	PC3	PC1	PC2	PC3
	213	67.8	62.7	211	57.1	59.0
	227	61.1	54.4	208	54.7	56.4
	222	68.3	55.9	217	60.6	55.9
	223	63.2	58.4	210	66.1	55.7
	230	62.4	58.6	226	57.1	59.3
	225	63.3	59.8	211	59.6	51.2
	225	61.8	55.9			
	221	66.4	58.1			
	221	64.9	57.0			
	206	58.2	55.9			
Mean	221.40	63.73	57.68	213.90	59.19	56.24

(a) Give interpretations or names to the first three principal components.

(b) Use the component scores to form linear discriminant functions for distinguishing among the four wolf groups.

(c) Show that, apart from round-off error, the within-groups covariance matrix of the component scores is merely the diagonal matrix of the first three characteristic roots of the principal-components analysis.

25. Give the decomposition of the general Hotelling T^2 statistic $(\mathbf{y} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\mathbf{y} - \boldsymbol{\mu})$ into the sum of squared univariate t statistics computed from the principal component scores. Use the diagonalization $\mathbf{S}^{-1} = \mathbf{P}\mathbf{D}^{-1}\mathbf{P}'$, where $\mathbf{P}$ has the characteristic vectors of $\mathbf{S}$ for columns, and $\mathbf{D}$ is the diagonal matrix of the successive characteristic vectors of $\mathbf{S}$.

26. Extend the result of Exercise 25 to the case of a singular sample with $N < p$, using the expression for a generalized inverse given in Section 4.7 in terms of principal components.

27. $N = 12$ normal young males had their cerebral blood flow[4] (CBF) measured at $p = 18$ regions of interest (ROIs) in the left brain hemisphere under a baseline condition P and a verbal stimulus condition R. The R-P differences of log(CBF) at the eighteen ROIs were used as measures of verbal stimulation in the brain, and for the test of the hypothesis of zero means. The actual differences, their

[4] I am indebted to Dr. Martin Reivich for the original cerebral blood flow values.

covariance matrix, and all eleven principal components with non-zero characteristic roots will be omitted to conserve space. Instead, these summary statistics will be given:

Principal Component	Characteristic root	Percent total variance	Component scores Mean	SD	t
1	2.1296268	95.99	0.88	1.46	2.09
2	0.0317019	1.43	−0.02	0.18	−0.30
3	0.0253782	1.14	0.05	0.16	1.15
4	0.0105563	0.48	0.04	0.10	1.52
5	0.0066151	0.30	0.02	0.08	0.89
6	0.0062699	0.28	−0.04	0.08	−1.64
7	0.0037580	0.17	0.02	0.06	1.24
8	0.0020452	0.09	−0.01	0.05	−0.89
9	0.0013852	0.06	−0.02	0.04	−1.96
10	0.0009547	0.04	0.0279	0.0309	3.13
11	0.0003289	0.01	0.0289	0.0181	5.52

	ROI	Name	R - P Mean	PC1	PC10	PC11	a
1.	2	Anterior cingulate gyrus	0.225	0.29	−0.45	−0.41	−0.50
2.	4	Superior frontal gyrus	0.152	0.23	0.03	−0.21	−0.25
3.	6	Medial frontal gyrus	0.187	0.24	0.14	0.06	0.09
4.	10	Superior temporal gyrus	0.209	0.21	0.02	−0.02	0.05
5.	12	Supramarginal gyrus	0.199	0.18	−0.15	0.25	0.22
6.	16	Caudate nucleus	0.191	0.24	−0.31	0.33	0.15
7.	20	Thalamus	0.232	0.27	0.18	−0.14	−0.04
8.	24	Cuneus	0.178	0.21	0.14	0.07	0.05
9.	26	Posterior cingulate gyrus	0.267	0.25	0.04	0.09	0.16
10.	30	Medial temporal gyrus	0.206	0.22	−0.21	0.00	−0.12
11.	32	Inferior frontal gyrus	0.206	0.24	0.20	−0.45	−0.35
12.	36	Putamen	0.244	0.28	−0.27	0.38	0.30
13.	38	Parahippocampal gyrus	0.207	0.22	0.48	0.05	0.18
14.	52	Lingual gyrus	0.216	0.24	−0.03	−0.21	−0.12
15.	68	Angular gyrus	0.209	0.20	0.19	0.20	0.27
16.	80	Inferior temporal gyrus	0.203	0.22	−0.04	−0.26	−0.33
17.	84	Fusiformal gyrus	0.217	0.24	0.38	0.25	0.34
18.	86	Transverse temporal gyrus	0.185	0.25	−0.14	0.11	0.04

(*a*) The linear function of the $p = 18$ differences giving the greatest t^2 statistic is defined by expression (5) of Section 4.7 in terms of the generalized inverse of the sample covariance matrix of the differences. The vector **a** normalized to unit length is shown in the last column of the second table. Compare the elements of **a** with those of the last two characteristic vectors PC10 and PC11.

(*b*) Find the greatest t^2 from the individual t statistics of the $r = 11$ principal components with positive characteristic roots.

28. Birren and Morrison (1961) analyzed a sample of $N = 933$ Wechsler Adult Intelligence Scale (WAIS) scores and the covariates of age and years of formal education of the test participants. The correlations of the eleven subtest scores, age, and education (multiplied by 100 to conserve space) are shown in the first table. The principal component coefficients, again multiplied by 100 to save space, are given in the second table. The subtest full names are given in Example 6.5. Finally, the characteristic roots

of the original correlation matrix follow.

Correlations ($\times$ **100**)

	2	3	4	5	6	7	8	9	10	11	12	13
1. Inf	67	62	66	47	81	47	60	49	51	41	−07	66
2. Comp		54	60	39	72	40	54	45	49	38	−08	52
3. Arith			51	51	58	41	46	48	43	37	−08	49
4. Sim				41	68	49	56	50	50	41	−19	55
5. DgSp					45	45	42	39	42	31	−19	43
6. Vocab						49	57	46	52	40	−02	62
7. DgSm							50	50	52	46	−46	57
8. PicC								61	59	51	−28	48
9. BlkD									54	59	−32	44
10. PicA										46	−37	49
11. ObjA											−28	40
12. Age												−29

Principal Component Coefficients ($\times$ **100**)

Variable	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
1. Inf	32	28	04	−13	07	02	−05	−18	17	05	10	−57	−62
2. Comp	29	27	−08	−20	−19	−11	−06	55	16	−49	−29	27	−16
3. Arith	28	21	10	42	12	−68	−21	−02	−18	−03	36	11	08
4. Sim	30	13	−01	−24	−05	−08	62	19	−21	55	09	23	−06
5. DgSp	24	−01	44	68	−15	37	17	16	22	07	−13	−01	00
6. Vocab	32	32	02	−19	02	15	−01	03	−08	−06	−02	−45	73
7. DgSm	28	−30	29	−09	28	21	13	−11	−61	−45	03	05	−14
8. PicC	30	−09	−27	−01	−36	15	21	−43	29	−30	45	25	08
9. BlkD	28	−22	−40	22	02	−26	17	−32	−03	02	−67	−10	03
10. PicA	28	−19	−05	−06	−54	15	−58	03	−34	32	−03	04	−06
11. ObjA	25	−25	−53	17	49	27	−15	39	10	11	24	−02	−01
12. Age	−13	67	−29	21	12	36	−08	−24	−31	02	−10	29	−09
13. Educ	29	01	33	−28	40	04	−30	−29	36	21	−18	42	08

Characteristic roots

Component	**Characteristic root**
1	6.696
2	1.414
3	0.798
4	0.710
5	0.553
6	0.462
7	0.436
8	0.402
9	0.386
10	0.372
11	0.340
12	0.262
13	0.170

(a) Make a scree plot of the characteristic roots, and determine if the plot has drops in the sequence of the characteristic roots.

(b) The second component is highly correlated with age. Interpret that component as a contrast of verbal (Subtests 1 – 6) and performance (Subtests 7 – 11) scores.

(c) Which sets of components might be considered ill-defined because of their nearly equal characteristic roots?

(d) The Wechsler subtests were undoubtedly constructed to measure unique facets of cognition. Do any of the later components appear to be highly correlated with the individual subtests?

29. Instructor averages on the scales of an older course evaluation form were published at the end of the semester at a business school. Covariance matrices were computed for selected scales of the form for three departments, and principal components were extracted from each matrix. The first two components are given in the table. The original covariance matrices will be omitted.

	Department					
	Accounting		Finance		Management	
	PC1	PC2	PC1	PC2	PC1	PC2
Instructor						
Presentation ability	0.35	−0.27	0.40	−0.56	0.32	−0.17
Stimulate interest	0.44	0.00	0.43	−0.03	0.43	−0.08
Subject knowledge	0.24	−0.21	0.22	−0.39	0.25	−0.03
Interest in students	0.35	−0.31	0.29	0.40	0.34	−0.28
Objectivity/fairness	0.26	−0.22	0.24	0.25	0.26	−0.41
Overall evaluation	0.41	−0.18	0.41	−0.18	0.38	−0.13
Course						
Impact: New skills	0.25	0.28	0.27	0.14	0.26	0.21
Impact: General	0.29	0.34	0.29	0.15	0.29	0.31
Value of readings	0.19	0.48	0.18	0.36	0.21	0.44
Overall evaluation	0.27	0.40	0.30	0.14	0.33	0.18
Workload demands	0.06	0.36	0.10	0.29	0.09	0.58
Variance	3.449	0.518	2.152	0.346	1.912	0.457
Proportion of total variance	0.763	0.115	0.701	0.113	0.628	0.150
Sample size	17		27		21	

(a) What differences appear to exist among the principal components of the evaluation scores of the three departments?

(b) Use an approximate large-sample statistic to test the hypothesis that the first characteristic roots of the Accounting and Finance population covariance matrices are equal.

(c) Find the 95% confidence intervals for the three largest characteristic roots.

30. *(a)* From the cerebral metabolic rate covariance matrix of Exercise 32, Chapter 2, these principal component coefficients and characteristic roots were extracted by the APL SYMEIG command:

Region	Component 1	2	3	4
Left thalamus	0.4787	−0.4695	0.1017	0.7349
Right thalamus	0.5013	−0.5462	−0.0396	−0.6699
Left frontal eyefield	0.4906	0.4312	−0.7548	0.0603
Right frontal eyefield	0.5280	0.5435	0.6468	−0.0862
Characteristic root	8.7567	0.3039	0.0240	0.0193

Interpret the components in terms of contrasts of the brain regions.

(b) The four regions have this correlation matrix:

$$\begin{bmatrix} 1 & 0.990962 & 0.935901 & 0.928158 \\ & 1 & 0.930254 & 0.921620 \\ & & 1 & 0.989432 \\ & & & 1 \end{bmatrix}$$

These components were computed from the correlation matrix by the SYMEIG command:

Region	Component 1	2	3	4
Left thalamus	0.5009	−0.4743	0.1463	0.7091
Right thalamus	0.4993	−0.5237	−0.0923	−0.6840
Left frontal eyefield	0.5010	0.4657	−0.7217	0.1065
Right frontal eyefield	0.4988	0.5328	0.6703	−0.1343
Characteristic root	3.8482	0.1327	0.0103	0.0088

Compare these components with those found from the covariance matrix. Do you think variation in the sample could be explained by two principal dimensions?

31. The magazine *U. S. News and World Report*[5] gave rankings and other data for the $N = 25$ highest-ranked schools of business in an annual survey of higher education institutions. We shall consider only these variables:

1. Reputation rank by academics
2. Reputation rank by business CEOs
3. Student selectively rank
4. Placement success rank
5. Graduation ratio rank
6. Mean GMAT score
7. Acceptance rate
8. Percent employed three months after graduation
9. 1993 median starting salary

[5] March 21, 1994, pp. 80–81

These correlations (rounded to conserve space) were computed:

	1	2	3	4	5	6	7	8	9
1	1	0.74	0.73	0.78	0.38	−0.69	0.58	−0.40	−0.82
2		1	0.53	0.51	0.25	−0.57	0.31	−0.32	−0.72
3			1	0.52	0.10	−0.91	0.79	−0.17	−0.68
4				1	0.41	−0.49	0.54	−0.77	−0.78
5					1	−0.12	0.00	−0.52	−0.45
6						1	−0.67	0.17	0.72
7							1	−0.28	−0.55
8								1	0.46
9									1

These characteristic vectors and roots were extracted from the correlation matrix:

	Characteristic vector								
Variable	1	2	3	4	5	6	7	8	9
1	0.40	0.00	0.19	0.07	0.52	−0.15	−0.52	−0.27	−0.41
2	0.32	−0.01	0.58	0.46	−0.30	−0.42	0.10	0.07	0.26
3	0.36	−0.37	−0.07	−0.25	−0.16	0.02	−0.26	0.75	−0.09
4	0.37	0.27	−0.29	0.26	0.33	0.32	−0.10	0.12	0.64
5	0.18	0.55	0.22	−0.73	−0.02	−0.22	0.01	−0.01	0.19
6	−0.36	0.34	−0.07	0.22	0.44	−0.42	0.14	0.54	−0.16
7	0.31	−0.31	−0.53	−0.11	0.10	−0.58	0.34	−0.22	0.09
8	−0.24	−0.52	0.43	−0.25	0.50	0.02	0.12	0.01	0.40
9	−0.40	−0.06	−0.18	0.01	−0.20	−0.37	−0.70	−0.09	0.36
Root	5.29	1.61	0.79	0.55	0.27	0.23	0.17	0.06	0.03

(*a*) Make a scree plot of the characteristic roots, and determine which components might be dropped from the analysis.

(*b*) Give interpretations and names to the principal components retained in part (*a*).

(*c*) Could all correlations be made positive by reversing the scales of variables 6, 8, and 9?