Chapter 4 *Multidimensional quantitative data*

4.0 Multidimensional statistics

Basic statistics are now part of the training of most ecologists. However, statistical techniques based on simple distributions such as the unidimensional normal distribution are not really appropriate for analysing complex ecological data sets. Nevertheless, researchers sometimes perform series of simple analyses on the various descriptors in their data set, expecting to obtain results that are pertinent to the problem under study. This type of approach is incorrect because it does not take into account the covariances among descriptors; see also Box 1.3 where the statistical problem created by multiple testing is explained. In addition, such an approach only extracts minimum information from data that have often been collected at great cost and it usually generates a mass of results from which it may be difficult to draw synthetic conclusions. Finally, in studies involving species assemblages, it is usually more interesting to describe the variability of the structure of the assemblage as a whole (i.e. *mensurative* variation observed through space or time, or *manipulative* variation resulting from experimental manipulation; Hurlbert, 1984) than to analyse each species independently.

Fortunately, methods derived from *multidimensional statistics*, which are used throughout this book, are designed for analysing complex data sets. These methods take into account the co-varying nature of ecological data and can evidence the structures that underlie the data. The present chapter discusses the basic theory and characteristics of multidimensional data analysis. Mathematics are kept to a minimum, so that readers can easily reach a high level of understanding. Many approaches of practical interest are discussed, including several types of linear correlation with their statistical tests. It must be noted that this chapter is limited to linear statistics.

A number of excellent textbooks deal with detailed aspects of multidimensional statistics, for example Mardia *et al*. (1979), Muirhead (1982), Anderson (2003), and Hair *et al*. (2010). There are also several titles on specialized topics such as linear

Table 4.1 Numerical example of two species observed at four sampling sites. Figure 4.1 shows that each row of the data matrix may be construed as a vector, as defined in Section 2.4.

Sampling sites (objects)	Species (descriptors) 1	2	($p = 2$)
1	5	1	
2	3	2	
3	8	3	
4	6	4	
($n = 4$)			

models, linear regression, and time series analysis. None of these books specifically deals with ecological data, however.

Multidimensional Multivariate

Several authors use the term *multivariate* as an abbreviation for *multidimensional variate* (the latter term meaning *random variable*; Section 1.0). As an adjective, *multivariate* is interchangeable with *multidimensional*.

4.1 Multidimensional variables and dispersion matrix

As stated in Section 1.0, the present textbook deals with the analysis of *random variables*. Ecological data matrices have n rows and p columns (Section 2.1). Each row is a *vector* (Section 2.4) which is, statistically speaking, one realization of a p-dimensional random variable. When, for example, p species are observed at n sampling sites, the species are the p dimensions of a random variable "species" and each site provides one realization of this p-dimensional random variable.

To illustrate this concept, four sampling units with two species (Table 4.1) are plotted in a two-dimensional Euclidean space (Fig. 4.1). Vector "site 1" is the doublet (5,1). It is plotted in the same two-dimensional space as the three other vectors "site i". Each row of the data matrix is a two-dimensional vector, which is one realization of the (bivariate) random variable "species". The random variable "species" is said to be two-dimensional because the sampling units (objects) contain two species (descriptors), the two dimensions being species 1 and 2, respectively. The species descriptors of this example are the axes of the attribute space, or A-space (Fig. 7.2).

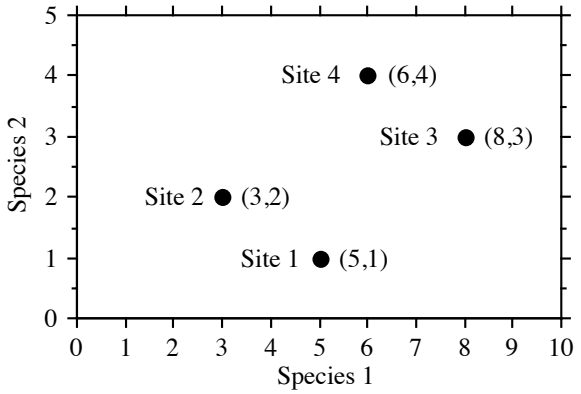


Figure 4.1 Four realizations (sampling sites from Table 4.1) of the two-dimensional random variable "species" are plotted in a two-dimensional Euclidean space.

As the number of descriptors (e.g. species) increases, the number of dimensions of the random variable "species" similarly increases, so that more axes are necessary to construct the space in which the objects are plotted. Thus, the p descriptors make up a p-dimensional random variable and the n vectors of observations (objects) are as many realizations of the p-dimensional vector "descriptors". The present chapter does not deal with *samples* of observations, which result from field or laboratory work (for a brief discussion on sampling, see Section 1.0). It focuses instead on *populations*, which are investigated by means of samples.

Before approaching the multidimensional normal distribution, it is necessary to define a p-dimensional random variable "descriptors":

$$\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots \mathbf{y}_j, \ldots \mathbf{y}_p] \qquad \textbf{(4.1)}$$

Each element $\mathbf{y}_j$ of the multidimensional variable $\mathbf{Y}$ is a one-dimensional random variable. Every descriptor $\mathbf{y}_j$ is observed in each of the n vectors "object", each sampling unit i providing a realization of the p-dimensional random variable.

In ecology, the structure of *dependence* among descriptors is, in many instances, the matter being investigated. Researchers who study multidimensional data using univariate statistics assume that the p unidimensional $\mathbf{y}_j$ variables in $\mathbf{Y}$ are *linearly independent* of one another (third meaning of *independence* in Box 1.1). This is the reason why univariate statistical methods are inappropriate with most ecological data and why methods that take into account the *dependence* among descriptors must be used to analyse multidimensional data sets. Only these methods will generate proper results when there is dependence among descriptors; it is never acceptable to replace a multidimensional analysis by a series of unidimensional treatments.

Table 4.2 Symbols used to identify (population) parameters and (sample) statistics.

	Parameter		Statistic	
	Matrix or vector	Elements	Matrix or vector	Elements
Covariance	$\boldsymbol{\Sigma}$ (sigma)	σ_{jk} (sigma)	$\mathbf{S}$	s_{jk}
Correlation	$\mathbf{P}$ (rho)	ρ_{jk} (rho)	$\mathbf{R}$	r_{jk}
Mean	$\boldsymbol{\mu}$ (mu)	μ_j (mu)	$\bar{\mathbf{y}}$	$\bar{y}_j$

The symbols for covariance matrix $\boldsymbol{\Sigma}$ and summation $\sum$ should not be confused.

The usual tests of significance require, however, "that successive sample observation vectors from the multidimensional population have been drawn in such a way that they can be construed as realizations of independent random vectors" (Morrison, 1990, p. 80). Subsection 1.1.1 has shown that this assumption of independence among observations is most often not realistic in ecology. Lack of independence among the observations (data rows) does not really matter when statistical models are used for descriptive purposes only, as it is often the case in the present book. For statistical testing, however, corrected tests of significance have to be used when the observations are spatially or temporally correlated (Subsection 1.1.2).

To sum up: (1) the p descriptors in ecological data matrices are the p dimensions of a random variable "descriptors"; (2) in general, the p descriptors are *not linearly independent* of one another; methods of multidimensional analysis are designed to bring out the structure of linear dependence among descriptors; (3) each of the n sampling units is a realization of the p-dimensional vector "descriptors"; (4) the usual tests of significance assume that the n sampling units are realizations of *independent* random vectors. The latter condition is generally not met in ecology, with consequences that were discussed in the previous paragraph and in Subsection 1.1.1. For the various meanings of the term *independence* in statistics, see Box 1.1.

Parameter
Statistic

Greek and roman letters are used here and in the remainder of the book (Table 4.2). The properties of a *population* (called *parameters*) are denoted by *greek* letters. Their *estimates* (called *statistics*), computed from *samples,* are symbolized by the corresponding *roman* letters. These conventions are complemented by those pertaining to matrix notation (Section 2.1).

The dependence among quantitative variables $\mathbf{y}_j$ brings up the concept of *covariance*. Covariance is the extension, to two descriptors, of the concept of *variance*. The variance is a measure of the *dispersion* of a random variable $\mathbf{y}_j$ around its mean; it is denoted σ_j^2. Covariance measures the *joint dispersion* of two random variables $\mathbf{y}_j$

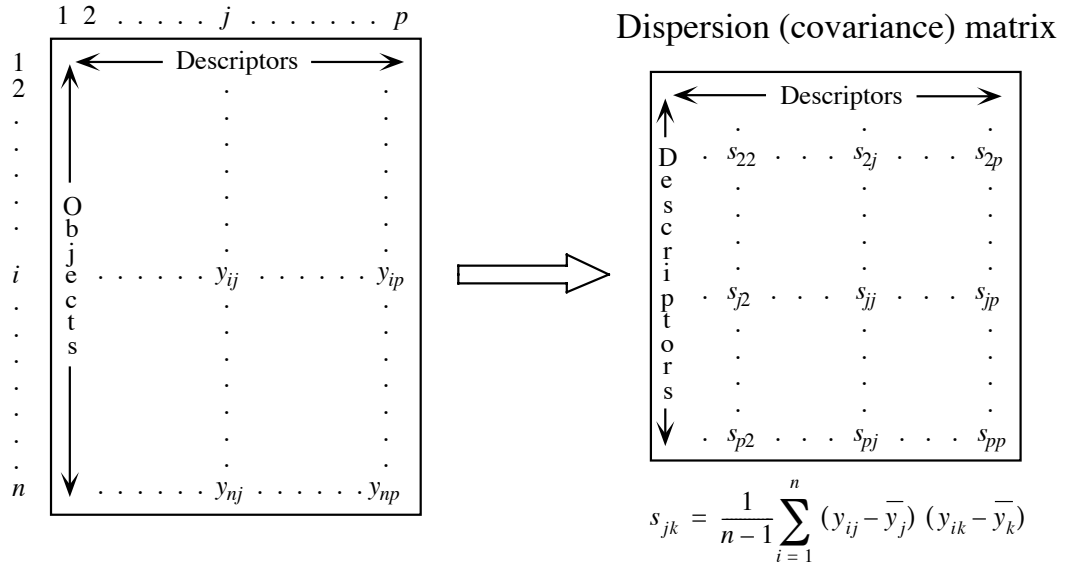


Figure 4.2 Structure of ecological data. Given their nature, ecological descriptors are, in most cases, *linearly dependent* on one another (Box 1.1).

Dispersion matrix

and $\mathbf{y}_k$ around their means; it is denoted σ_{jk}. The *dispersion matrix* of $\mathbf{Y}$, called matrix $\mathbf{\Sigma}$ (sigma), contains the variances and covariances of the p descriptors (Fig. 4.2):

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix} \qquad \textbf{(4.2)}$$

Matrix $\mathbf{\Sigma}$ is an *association matrix* [descriptors × descriptors] (Section 2.2). The elements σ_{jk} of matrix $\mathbf{\Sigma}$ are the covariances between all pairs of the p random variables. The matrix is symmetric because the covariance of $\mathbf{y}_j$ and $\mathbf{y}_k$ is identical to that of $\mathbf{y}_k$ and $\mathbf{y}_j$. Each diagonal element of $\mathbf{\Sigma}$ is the covariance of a descriptor $\mathbf{y}_j$ with itself, which is the variance of $\mathbf{y}_j$, so that $\sigma_{jj} = \sigma_j^2$.

Variance

The *estimate* of the *variance* of $\mathbf{y}_j$, denoted s_j^2, is computed on the *centred variable* $(y_{ij} - \bar{y}_j)$. Variable $\mathbf{y}_j$ is centred by subtracting the mean $\bar{y}_j$ from each of the n

observations y_{ij}. As a result, the mean of the centred variable is zero. The unbiased estimator of the population variance s_j^2 is computed using the well-known formula:

$$\text{var}(\mathbf{y}_j) = s_j^2 = \frac{1}{n-1}\sum_{i=1}^{n} (y_{ij} - \bar{y}_j)^2 \qquad \textbf{(4.3)}$$

where the sum of *squares* of the *centred data*, for descriptor j, is divided by the number of objects minus one $(n - 1)$. The summation is over the n observations of descriptor j. The variance of $\mathbf{y}_j$ is expressed in the squared physical dimension of $\mathbf{y}_j$. In the same way, the estimate (s_{jk}) of the *covariance* (σ_{jk}) of $\mathbf{y}_j$ and $\mathbf{y}_k$ is computed on the centred variables $(y_{ij} - \bar{y}_j)$ and $(y_{ik} - \bar{y}_k)$, using the formula of a "bivariate variance". The *covariance* s_{jk} is calculated as:

Covariance

$$\text{cov}(\mathbf{y}_j,\mathbf{y}_k) = s_{jk} = \frac{1}{n-1}\sum_{i=1}^{n} (y_{ij} - \bar{y}_j)\,(y_{ik} - \bar{y}_k) \qquad \textbf{(4.4)}$$

When $k = j$, eq. 4.4 is identical to eq. 4.3. The positive square root of the variance is called the *standard deviation* (σ_j); it has the same dimension as $\mathbf{y}_j$. Its estimate s_j is:

Standard deviation

$$s_j = \sqrt{s_j^2} \qquad \textbf{(4.5)}$$

The coefficient of variation is a dimensionless measure of variation. *CV* is used to compare the variation of variables expressed in different physical units. It is obtained by dividing the standard deviation s_j by the mean $\bar{x}_j$ of variable j:

Coefficient of variation

$$CV_j = s_j / \bar{x}_j$$

Since the standard deviation and the mean of a variable have the same physical units, CV_j is dimensionless. CV_j is only defined for quantitative variables that have non-zero means and it does not make sense for interval-scale variables (Subsection 1.4.1), for which the value of the mean is arbitrary. The coefficient of variation may be rescaled to percentages by multiplying its value by 100. For small n, an estimate with reduced bias is obtained by multiplying *CV* by $(1 + 1/(4n))$.

Contrary to the variance, which is always positive, the covariance may take positive or negative values. To understand the meaning of the covariance, imagine that the object points are plotted in a scatter diagram where the axes are descriptors $\mathbf{y}_j$ and $\mathbf{y}_k$. The data are centred by drawing new axes, whose origin is at the centroid $(\bar{y}_j, \bar{y}_k)$ of the cloud of points (centred plots of that kind with positive and negative correlations are shown in Fig. 4.7). A positive covariance (e.g. Fig. 4.7, right) means that most of the points are in quadrants I and III of the centred plot, where the centred values $(y_{ij} - \bar{y}_j)$ and $(y_{ik} - \bar{y}_k)$ have the same signs. This corresponds to a positive relationship between the two descriptors. The converse is true for a negative covariance (e.g. Fig. 4.7, left), for which most of the points are in quadrants II and IV

of the centred plot. When the covariance is null (e.g. Fig. 4.8, left) or small, the points are equally distributed among the four quadrants of the centred plot.

The covariance or dispersion matrix* **S** can be computed directly by multiplying the *matrix of centred data* $[y - \bar{y}]$ with its transpose $[y - \bar{y}]'$:

$$\text{cov}(\mathbf{Y}) = \mathbf{S} = \frac{1}{n-1}[y - \bar{y}]'[y - \bar{y}] \qquad \textbf{(4.6)}$$

$$\mathbf{S} = \frac{1}{n-1}\begin{bmatrix} (y_{11}-\bar{y_1}) & (y_{21}-\bar{y_1}) & \cdots & (y_{n1}-\bar{y_1}) \\ (y_{12}-\bar{y_2}) & (y_{22}-\bar{y_2}) & \cdots & (y_{n2}-\bar{y_2}) \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ (y_{1p}-\bar{y_p}) & (y_{2p}-\bar{y_p}) & \cdots & (y_{np}-\bar{y_p}) \end{bmatrix} \begin{bmatrix} (y_{11}-\bar{y_1}) & (y_{12}-\bar{y_2}) & \cdots & (y_{1p}-\bar{y_p}) \\ (y_{21}-\bar{y_1}) & (y_{22}-\bar{y_2}) & \cdots & (y_{2p}-\bar{y_p}) \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ (y_{n1}-\bar{y_1}) & (y_{n2}-\bar{y_2}) & \cdots & (y_{np}-\bar{y_p}) \end{bmatrix}$$

$$\mathbf{S} = \frac{1}{n-1}\begin{bmatrix} \sum_{i=1}^{n}(y_{i1}-\bar{y_1})^2 & \sum_{i=1}^{n}(y_{i1}-\bar{y_1})(y_{i2}-\bar{y_2}) & \cdots & \sum_{i=1}^{n}(y_{i1}-\bar{y_1})(y_{ip}-\bar{y_p}) \\ \sum_{i=1}^{n}(y_{i2}-\bar{y_2})(y_{i1}-\bar{y_1}) & \sum_{i=1}^{n}(y_{i2}-\bar{y_2})^2 & \cdots & \sum_{i=1}^{n}(y_{i2}-\bar{y_2})(y_{ip}-\bar{y_p}) \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \sum_{i=1}^{n}(y_{ip}-\bar{y_p})(y_{i1}-\bar{y_1}) & \sum_{i=1}^{n}(y_{ip}-\bar{y_p})(y_{i2}-\bar{y_2}) & \cdots & \sum_{i=1}^{n}(y_{ip}-\bar{y_p})^2 \end{bmatrix}$$

This elegant and rapid procedure shows once again the advantage of matrix algebra in numerical ecology, where the data sets are generally large.

Numerical example. Four species ($p = 4$) were observed at five stations ($n = 5$). The estimated population parameters, for the species, are the means ($\bar{y}_j$), the variances (s_j^2), and the covariances (s_{jk}). The original and centred data are shown in Table 4.3. Because $s_{jk} = s_{kj}$, the dispersion matrix is symmetric. The mean of each *centred variable* is zero.

In this numerical example, the covariance between species 2 and the other three species is zero. This does not necessarily mean that species 2 is independent of the other three, but simply that the joint *linear* dispersion of species 2 with any one of the other three is zero. This example will be revisited in Section 4.2.

* Some authors call $[y - \bar{y}]'[y - \bar{y}]$ a *dispersion matrix* and **S** a *covariance matrix*. For these authors, a covariance matrix is then a dispersion matrix divided by $(n - 1)$.

Table 4.3 Numerical example. Calculation of centred data and covariances.

Sites	Original data	Centred data
1 2 3 4 5	$\mathbf{Y} = \begin{bmatrix} 1 & 5 & 2 & 6 \\ 2 & 2 & 1 & 8 \\ 3 & 1 & 3 & 4 \\ 4 & 2 & 5 & 0 \\ 5 & 5 & 4 & 2 \end{bmatrix}$	$[y - \bar{y}] = \begin{bmatrix} -2 & 2 & -1 & 2 \\ -1 & -1 & -2 & 4 \\ 0 & -2 & 0 & 0 \\ 1 & -1 & 2 & -4 \\ 2 & 2 & 1 & -2 \end{bmatrix}$
Means	$\bar{\mathbf{y}}' = \begin{bmatrix} 3 & 3 & 3 & 4 \end{bmatrix}$	$[\overline{y - \bar{y}}]' = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}$
$n - 1 = 4$	$\mathbf{S} = \frac{1}{n-1}[y - \bar{y}]'[y - \bar{y}] = \begin{bmatrix} 2.5 & 0 & 2 & -4 \\ 0 & 3.5 & 0 & 0 \\ 2 & 0 & 2.5 & -5 \\ -4 & 0 & -5 & 10 \end{bmatrix}$	

The square root of the determinant of the dispersion matrix $|\mathbf{S}|^{1/2}$ is known as the *generalized variance*. It is also equal to the square root of the product of the eigenvalues of **S**.

Any dispersion matrix **S** is *positive semidefinite* (Table 2.2). Indeed, the quadratic form of **S** ($p \times p$) with any real and non-null vector **t** (of size p) is:

$$\mathbf{t}'\mathbf{S}\mathbf{t}$$

This expression can be expanded using eq. 4.6:

$$\mathbf{t}'\mathbf{S}\mathbf{t} = \mathbf{t}'\frac{1}{n-1}[y - \bar{y}]'[y - \bar{y}]\mathbf{t}$$

$$\mathbf{t}'\mathbf{S}\mathbf{t} = \frac{1}{n-1}[[y - \bar{y}]\mathbf{t}]'[[y - \bar{y}]\mathbf{t}] = \text{a scalar}$$

This scalar is the variance of the variable resulting from the product **Yt**. Since a variance, which is a sum of squared values, can only be positive or null, it follows that:

$$\mathbf{t}'\mathbf{S}\mathbf{t} \geq 0$$

so that **S** is positive semidefinite. This means that **S** cannot have negative eigenvalues.

This important property can be derived by computing the quadratic form of the dispersion matrix **S** using eq. 2.28 (right), $\mathbf{\Lambda} = \mathbf{U}^{-1}\mathbf{A}\mathbf{U}$. Because **S** is symmetric, its eigenvectors found in matrix **U** are orthogonal. Since they are also normalized, **U** is an orthonormal matrix, hence $\mathbf{U}^{-1} = \mathbf{U}'$ (property #7 of inverses, Section 2.8), and eq. 2.28 (right) can be written:

$$\mathbf{U}'\mathbf{S}\mathbf{U} = \mathbf{\Lambda}$$

In the quadratic form, vector **t** is replaced by each successive eigenvector $\mathbf{u}_j$ in turn, i.e. each column of matrix **U**. For each vector $\mathbf{u}_j$, the development above shows that

$$\mathbf{u}_j'\mathbf{S}\mathbf{u}_j \geq 0$$

Since $\mathbf{u}_j'\mathbf{S}\mathbf{u}_j = \lambda_j$, this demonstrates that *all eigenvalues* λ_j *of* **S** *are positive or null*. This property of dispersion matrices is fundamental in numerical ecology: it allows one to partition the variance of a matrix **Y** among real (i.e. non-imaginary) *principal axes* (Sections 4.4 and 9.1).

Another property of the dispersion matrix is that the sum of all values in **S** is equal to the variance of a synthetic variable **y** computed as the sum by rows (objects) of all descriptors in **Y**. For example, if **Y** contains species abundance data, the sum by rows (sites) of all species abundances is a new variable **y** corresponding to the total number of individuals at the sites, which can in some cases be interpreted as the total yield or the support capacity of the sites. If **Y** consists of species presence-absence data, **y** is the species richness of the sites. The variance of the synthetic variable **y** can be obtained by summing all values in **S** instead of computing **y** and then its variance. This property will be used in Subsection 13.1.4.

Ideally, matrix **S** (of order p) should be estimated from a number of observations n larger than the number of descriptors p. When $n \leq p$, the rank of matrix **S** is $n - 1$ and, consequently, only $n - 1$ of its rows or columns are independent of one another, so that $p - (n - 1)$ null eigenvalues are produced. The only practical consequence of $n \leq p$ is thus the presence of null eigenvalues in the principal component solution (Section 9.1). The first few eigenvalues of **S**, which are generally those of interest, have positive eigenvalues.

4.2 Correlation matrix

The previous section has shown that the covariance provides information on the orientation of the cloud of data points in the space defined by the descriptors. That statistic, however, does not provide any information on the intensity of the relationship between variables $\mathbf{y}_j$ and $\mathbf{y}_k$. Indeed, the covariance may increase or decrease without changing the relationship between $\mathbf{y}_j$ and $\mathbf{y}_k$. For example, in Fig. 4.3, the two clouds of points correspond to different covariance values (factor two in size, and thus in

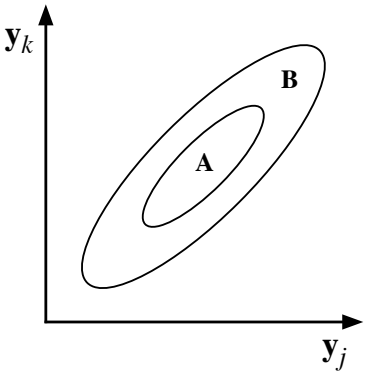


Figure 4.3 Several observations (objects), with descriptors $\mathbf{y}_j$ and $\mathbf{y}_k$, were made under two different sets of conditions (A and B). The two ellipses delineate clouds of point-objects corresponding to A and B, respectively. The covariance of $\mathbf{y}_j$ and $\mathbf{y}_k$ is twice as large for B as it is for A (larger ellipse), but the correlation between the two descriptors is the same in these two cases (i.e. the ellipses have the same shape).

covariance), but the relationship between the variables is identical (same shape). Since the covariance depends on the dispersion of the points around the mean of each variable (i.e. their variances), determining the intensity of the relationship between variables requires to control for the variances.

The *covariance* measures the joint *dispersion* of two random variables around the bivariate mean. The *correlation* is defined as a measure of the *dependence* between two random variables $\mathbf{y}_j$ and $\mathbf{y}_k$. As explained in Section 1.5, it often happens that matrices of ecological data contain descriptors with scales that are not commensurate, e.g. when some species have larger biomass than others by orders of magnitude, or when the descriptors have different physical dimensions (Chapter 3). Calculating covariances on such variables obviously does not make sense, except if the descriptors are first reduced to a common scale. The standardization procedure consists in centring all descriptors on a zero mean and reducing them to unit standard deviation (eq. 1.12). By using *standardized descriptors*, it is possible to calculate meaningful covariances because the new variables have the same scale (i.e. unit standard deviation) and are dimensionless (see Chapter 3).

Linear correlation

The covariance of two standardized descriptors is called the coefficient of linear correlation (Pearson *r*). This statistic has been proposed by the statistician Karl Pearson and is named after him. Given two standardized descriptors (eq. 1.12)

$$z_{ij} = \frac{y_{ij} - \overline{y}_j}{s_j} \qquad \text{and} \qquad z_{ik} = \frac{y_{ik} - \overline{y}_k}{s_k}$$

calculating their covariance (eq. 4.4) gives

$$s(z_j,z_k) = \frac{1}{n-1}\sum_{i=1}^{n} (z_{ij} - 0)\,(z_{ik} - 0) \qquad \text{because} \qquad \overline{z}_j = \overline{z}_k = 0$$

$$s(z_j,z_k) = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{y_{ij}-\overline{y_j}}{s_j}\right)\left(\frac{y_{ik}-\overline{y_k}}{s_k}\right)$$

$$s(z_j,z_k) = \left(\frac{1}{s_j s_k}\right)\frac{1}{n-1}\sum_{i=1}^{n}(y_{ij}-\overline{y_j})\,(y_{ik}-\overline{y_k})$$

$$s(z_j,z_k) = \left(\frac{1}{s_j s_k}\right)s_{jk} = r_{jk}\text{, the } \textit{coefficient of linear correlation} \text{ between } \mathbf{y}_j \text{ and } \mathbf{y}_k.$$

The developed formula is:

$$\text{cor}(\mathbf{y}_j,\mathbf{y}_k) = r_{jk} = \frac{s_{jk}}{s_j s_k} = \frac{\sum_{i=1}^{n}(y_{ij}-\overline{y_j})\,(y_{ik}-\overline{y_k})}{\sqrt{\sum_{i=1}^{n}(y_{ij}-\overline{y_j})^2\sum_{i=1}^{n}(y_{ik}-\overline{y_k})^2}} \qquad \textbf{(4.7)}$$

Correlation matrix

As in the case of dispersion (Section 4.1), it is possible to construct the *correlation matrix* of $\mathbf{Y}$, i.e. the $\mathbf{P}$ (rho) matrix, whose elements are the coefficients of linear correlation ρ_{jk}:

$$\mathbf{P} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{bmatrix} \qquad \textbf{(4.8)}$$

The *correlation matrix is the dispersion matrix of the standardized variables*. This concept will play a fundamental role in principal component analysis (Section 9.1). It should be noted that the diagonal elements of $\mathbf{P}$ are all equal to 1. This is because the comparison of any descriptor with itself is a case of complete dependence, which leads to a correlation $\rho = 1$. When $\mathbf{y}_j$ and $\mathbf{y}_k$ are independent of each other, $\rho_j = 0$. However, a correlation equal to zero does not necessarily imply that $\mathbf{y}_j$ and $\mathbf{y}_k$ are independent of each other, as shown by the following numerical example. A correlation $\rho_{jk} = -1$ is indicative of a complete, but inverse dependence of the two variables.

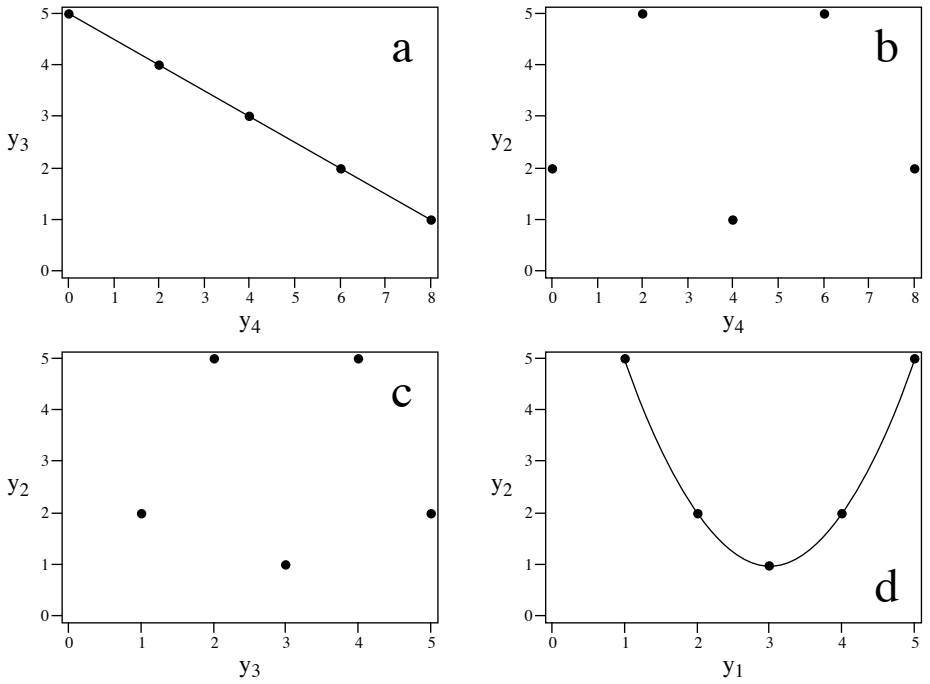


Figure 4.4 Numerical example. Relationships between species (a) 3 and 4, (b) 2 and 4, (c) 2 and 3, and (d) 2 and 1.

Numerical example. Using the values in Table 4.3, matrix **R** can easily be computed. According to eq. 4.7, each element r_{jk} combines the covariance s_{jk} with the variances s_j and s_k:

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0.8 & -0.8 \\ 0 & 1 & 0 & 0 \\ 0.8 & 0 & 1 & -1 \\ -0.8 & 0 & -1 & 1 \end{bmatrix}$$

Matrix **R** is symmetric, as was matrix **S**. The correlation $r = -1$ between species 3 and 4 means that these species are fully, but inversely, dependent (Fig. 4.4a). Correlations $r = 0.8$ and -0.8 are interpreted as indications of strong dependence between species 1 and 3 (direct) and species 1 and 4 (inverse), respectively. The *zero* correlation between species 2 and the other three species must be interpreted with caution. Figure 4.4d clearly shows that species 1 and 2 are completely *dependent* on each other since they are related by equation $y_2 = 1 + (3 - y_1)^2$; the zero correlation is, in this case, a consequence of the *linear* model underlying statistic *r*. Therefore, only the correlations that are *significantly* different from zero should be considered, since a null correlation has no unique interpretation.

Table 4.4 Numerical example. Calculation of standardized data and correlations.

Sites	Original data	Standardized data
1	$\mathbf{Y} = \begin{bmatrix} 1 & 5 & 2 & 6 \\ 2 & 2 & 1 & 8 \\ 3 & 1 & 3 & 4 \\ 4 & 2 & 5 & 0 \\ 5 & 5 & 4 & 2 \end{bmatrix}$	$\mathbf{Z} = \begin{bmatrix} -1.27 & 1.07 & -0.63 & 0.63 \\ -0.63 & -0.53 & -1.27 & 1.27 \\ 0 & -1.07 & 0 & 0 \\ 0.63 & -0.53 & 1.27 & -1.27 \\ 1.27 & 1.07 & 0.63 & -0.63 \end{bmatrix}$
2		
3		
4		
5		
Means	$\bar{\mathbf{y}}' = \begin{bmatrix} 3 & 3 & 3 & 4 \end{bmatrix}$	$\bar{\mathbf{z}}' = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}$
$n - 1 = 4$	$\mathbf{R}(y) = \mathbf{S}(z) = \frac{1}{n-1}\mathbf{Z}'\mathbf{Z} = \begin{bmatrix} 1 & 0 & 0.8 & -0.8 \\ 0 & 1 & 0 & 0 \\ 0.8 & 0 & 1 & -1 \\ -0.8 & 0 & -1 & 1 \end{bmatrix}$	

Since the correlation matrix is the dispersion matrix of standardized variables, it is possible, as in the case of matrix **S** (eq. 4.6), to compute **R** directly by multiplying the *matrix of standardized data* with its transpose:

$$\text{cor}(\mathbf{Y}) = \mathbf{R} = \frac{1}{n-1}\left[(y - \bar{y}) / s_y\right]'\left[(y - \bar{y}) / s_y\right] = \frac{1}{n-1}\mathbf{Z}'\mathbf{Z} \qquad \textbf{(4.9)}$$

Table 4.4 shows how to calculate correlations r_{jk} of the example as in Table 4.3, using this time the *standardized data*. The mean of each *standardized variable is* zero and its standard deviation is equal to unity. The *dispersion* matrix of **Z** is identical to the *correlation* matrix of **Y**, which was calculated above using the covariances and variances.

Matrices $\mathbf{\Sigma}$ and $\mathbf{P}$ are related to each other by the diagonal matrix of standard deviations of $\mathbf{Y}$. This new matrix, which was specifically designed here to relate $\mathbf{\Sigma}$ and $\mathbf{P}$, is symbolized by $\mathbf{D}(\sigma)$ and its inverse by $\mathbf{D}(\sigma)^{-1}$:

$$\mathbf{D}(\sigma) = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ 0 & \cdot & \dots & \sigma_p \end{bmatrix} \quad \text{and} \quad \mathbf{D}(\sigma)^{-1} = \begin{bmatrix} 1/\sigma_1 & 0 & \dots & 0 \\ 0 & 1/\sigma_2 & \dots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ 0 & \cdot & \dots & 1/\sigma_p \end{bmatrix}$$

Using these two matrices, one can write:

$$\mathbf{P} = \mathbf{D}(\sigma^2)^{-1/2}\ \mathbf{\Sigma}\ \mathbf{D}(\sigma^2)^{-1/2} = \mathbf{D}(\sigma)^{-1}\ \mathbf{\Sigma}\ \mathbf{D}(\sigma)^{-1} \qquad \textbf{(4.10)}$$

where $\mathbf{D}(\sigma^2)$ is the matrix of the diagonal elements of $\mathbf{\Sigma}$. It follows from eq. 4.10 that:

$$\mathbf{\Sigma} = \mathbf{D}(\sigma)\ \mathbf{P}\ \mathbf{D}(\sigma) \qquad \textbf{(4.11)}$$

Significance of *r*

The theory underlying tests of significance is summarized in Section 1.2. In the case of *r*, inference about the statistical population is in most instances through the null hypothesis H_0: $\rho = 0$. H_0 may also state that ρ has some other value than zero, which would be derived from ecological hypotheses. The general formula for testing correlation coefficients is given in Section 4.5 (eq. 4.39). The Pearson correlation coefficient r_{jk} involves two descriptors $\mathbf{y}_j$ and $\mathbf{y}_k$ (hence $m = 1$ when testing a coefficient of simple linear correlation using eq. 4.39), so that $\nu_1 = 1$ and $\nu_2 = n - 2 = \nu$. The general formula then becomes:

$$F = \frac{r_{jk}^2 / 1}{(1 - r_{jk}^2)/\nu} = \nu \frac{r_{jk}^2}{1 - r_{jk}^2} \qquad \textbf{(4.12)}$$

where $\nu = n - 2$. Statistic F is tested against $F_{\alpha[1,\nu]}$. Since the square root of a statistic $F_{[\nu_1,\nu_2]}$ is a statistic $t_{[\nu = \nu_2]}$ when $\nu_1 = 1$, r may also be tested using:

$$t = \frac{r_{jk}\sqrt{\nu}}{\sqrt{1 - r_{jk}^2}} \qquad \textbf{(4.13)}$$

The t-statistic is tested against the value $t_{\alpha[\nu]}$. In other words, H_0 is tested by comparing the F (or t) statistic to the value found in a table of critical values of F (or t). Equations 4.12 and 4.13 produce identical tests. The number of degrees of freedom is $\nu = (n - 2)$ because calculating a correlation coefficient requires prior estimation of two parameters, i.e. the means of the two populations (eq. 4.7). H_0 is rejected when the probability corresponding to F (or t) is smaller than or equal to a predetermined *level*

of significance (α for a two-tailed test, and $\alpha/2$ for a one-tailed test; the difference between the two types of tests is explained in Section 1.2). In principle, this test requires that the sample of observations be drawn from a population having a *bivariate normal distributio*n (Section 4.3). Testing for normality and multinormality is discussed in Section 4.6, and normalizing transformations in Section 1.5. When the data do not satisfy the condition of normality, t can be tested by permutation, as explained in Section 1.2.

Test of independence of variables

It is also possible to test the *independence of all variables* in a data matrix by considering the set of all correlation coefficients found in matrix **R**. The null hypothesis here is that the $p(p-1)/2$ coefficients are all equal to zero, H_0: $\mathbf{R} = \mathbf{I}$ (unit matrix). According to Bartlett (1954), the determinant of **R**, $|\mathbf{R}|$, can be transformed into a X^2 (chi-square) test statistic:

$$X^2 = -[n - (2p + 5)/6] \log_e |\mathbf{R}| \qquad \textbf{(4.14)}$$

where $\log_e |\mathbf{R}|$ is the natural logarithm of the determinant of **R**. This statistic is approximately distributed as χ^2 with $\nu = p(p-1)/2$ degrees of freedom. When the probability associated with X^2 is significantly low, the null hypothesis of complete independence of the p descriptors is rejected. In principle, this test requires the observations to be drawn from a population with a *multivariate normal distributio*n (Section 4.3). If the null hypothesis of independence of all variables is rejected, the $p(p-1)/2$ correlation coefficients in matrix **R** may be tested individually. Box 1.3 describes how to correct individual p-values in situations of multiple testing.

Other correlation coefficients are described in Sections 4.5 and 5.3. When the coefficient of linear correlation must be distinguished from other coefficients, it is referred to as *Pearson r*. Elsewhere, r is called the *coefficient of linear correlation* or *correlation coefficient*. Table 4.5 summarizes the main properties of this coefficient.

4.3 Multinormal distribution

In general, the mathematics of the normal distribution is of little concern to ecologists using unidimensional statistical methods. In the best case, data are normalized (Section 1.5) before being subjected to tests that are based on *parametric* hypotheses. It must be remembered that all *parametric tests* require the data to follow a specific *distribution,* most often the normal distribution. When the data do not obey this condition, the results of parametric tests may be *invalid*.

There also exist nonparametric tests (Chapter 5), in which no reference is made to any theoretical distribution of the population, hence no use of parameters. That is also the case with permutation tests based on the usual parametric statistics, e.g. the Pearson correlation coefficient r (Subsection 1.2.2). Another advantage of nonparametric and permutational tests is that they remain valid for samples of very

Table 4.5 Main properties of the coefficient of linear correlation. Some of these properties are discussed in a later sections.

	Properties	Sections
1.	The coefficient of linear correlation measures the *intensity of the linear relationship* between two random variables.	4.2
2.	The coefficient of linear correlation between two variables can be calculated using their respective *variances* and their *covariance*.	4.2
3.	The correlation matrix is the *dispersion* matrix of *standardized variables*.	4.2
4.	The square of the coefficient of linear correlation is the *coefficient of determination*. It measures how much of the variance of each variable is explained by the other.	10.3
5.	The coefficient of linear correlation is a *parameter* of a multinormal distribution.	4.3
6.	The absolute value of the coefficient of linear correlation is the *geometric mean* of the *coefficients of linear regression* of each variable on the other.	10.3

small sizes, which are often encountered in ecological research. These tests are of great interest to ecologists. Researchers may nevertheless attempt to normalize their data to have access to the powerful toolbox of parametric statistics or because some of the methods of multivariate analysis, e.g. principal component analysis (Section 9.1), perform better when the response data distributions are not strongly asymmetric.

Multidimensional statistics require careful examination of the main characteristics of the *multinormal* (or *multivariate normal*) *distribution*. Several of the methods described in the present chapter, and also in Chapters 9, 10 and 11, are founded on principles derived from the multinormal distribution. This is true even in cases where no test of significance is performed, which is often the case in numerical ecology (i.e. descriptive versus inferential statistics, Sections 1.2).

The logic of an approach centred on the multinormal distribution is based upon a theorem which is undoubtedly one of the most important of statistics. According to the *central limit theorem*, when a random variable results from several independent and additive effects, of which none has a dominant variance, then this variable tends towards a normal distribution even if the effects are not themselves normally distributed. Since ecological variables, and species abundances in particular, are often influenced by several independent random factors, the above theorem explains why the normal distribution is frequently invoked to describe ecological phenomena. This justifies a careful examination of the properties of the multinormal distribution before studying the methods for analysing multidimensional quantitative data.

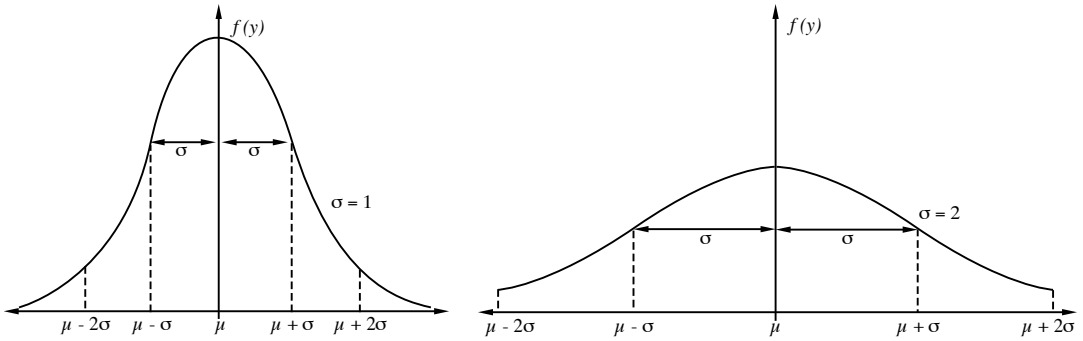


Figure 4.5 Role of the standard deviation σ in the normal distribution function. The abscissa is variable *y*.

Normal

The probability density of a *normal* random variable *y* is:

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right] \qquad \textbf{(4.15)}$$

(Laplace-Gauss equation) where exp […] reads "*e* to the power […]", *e* being the Napierian base (*e* = 2.71828…). Calculation of $f(y)$, for a given value *y*, only requires μ and σ. The mean (μ) and standard deviation (σ) of the theoretical population completely determine the shape of the probability distribution. This is why they are called the *parameters* of the normal distribution. The curve is symmetric on both sides of μ and its exact shape depends on σ (Fig. 4.5).

The value σ determines the positions of the inflexion points along the normal curve. These points are located on both sides of μ, at a distance σ, whereas μ positions the curve on the abscissa. In Fig. 4.5, the surface under each of the two curves is identical for the same number of σ units on either side of μ. The height of the curve is the probability density corresponding to the *y* value; for a continuous function such as that of the normal distribution, the probability of finding a value between $y = a$ and $y = b$ ($a < b$) is given by the surface under the curve between *a* and *b*. For example, the probability of finding a value between $\mu - 1.96\sigma$ and $\mu + 1.96\sigma$ is 0.95.

In view of examining the properties of the multinormal distribution, it is convenient to first consider the joint probability density of *p* *independent* unidimensional normal variables. For *each* of these *p* variables y_j, the probability density is given by eq. 4.15, with mean μ_j and standard deviation σ_j:

$$f(y_j) = \frac{1}{\sqrt{2\pi}\sigma_j}\exp\left[-\frac{1}{2}\left(\frac{y_j-\mu_j}{\sigma_j}\right)^2\right] \qquad \textbf{(4.16)}$$

A basic law of probabilities states that the joint probability density of several independent variables is the product of their individual densities. It follows that the joint probability density for p independent variables is:

$$f(y_1, y_2, \ldots, y_p) = f(y_1) \times f(y_2) \times \ldots \times f(y_p)$$

$$f(y_1, y_2, \ldots, y_p) = \frac{1}{(2\pi)^{p/2}\sigma_1\sigma_2\ldots\sigma_p} \exp\left[-\frac{1}{2} \sum_{j=1}^{p} \left(\frac{y_j - \mu_j}{\sigma_j} \right)^2 \right] \qquad \textbf{(4.17)}$$

Using the conventions of Table 4.2, one defines the following matrices:

$$\mathbf{y} = \left[y_1 \; y_2 \; \ldots \; y_p \right] \qquad \boldsymbol{\mu} = \left[\mu_1 \; \mu_2 \; \ldots \; \mu_p \right] \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & \ldots & 0 \\ \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \ldots & \cdot \\ 0 & 0 & \ldots & \sigma_p^2 \end{bmatrix} \qquad \textbf{(4.18)}$$

where $\mathbf{y} = [y_1 \; y_2 \; \ldots \; y_p]$ is the p-dimensional vector of coordinates of a point for which the probability density, i.e. the height (ordinate) of the p-dimensional normal curve, is sought; $\boldsymbol{\mu}$ is the vector of means, and $\boldsymbol{\Sigma}$ is the dispersion matrix among the p independent variables. The determinant of $\boldsymbol{\Sigma}$ is the *generalized variance* of the multivariate distribution. The determinant of a diagonal matrix being equal to the product of the diagonal elements (Section 2.6), it follows that:

Generalized variance

$$\left| \boldsymbol{\Sigma} \right|^{1/2} = (\sigma_1 \, \sigma_2 \ldots \sigma_p)$$

From definitions (4.18), and for a single row vector $[y - \mu]$ of p-dimensional centred data, one can write:

$$[y - \mu] \, \boldsymbol{\Sigma}^{-1} \, [y - \mu]' = \sum_{j=1}^{p} \left(\frac{y_j - \mu_j}{\sigma_j} \right)^2$$

which is a scalar. Do not confuse, here, the summation symbol $\sum$ with dispersion matrix $\boldsymbol{\Sigma}$. Using these relationships, eq. 4.17 is rewritten as:

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} \exp \{ -(1/2) \, [y - \mu] \, \boldsymbol{\Sigma}^{-1} \, [y - \mu]' \} \qquad \textbf{(4.19)}$$

The above equations are for the joint probability density of p *independent* unidimensional normal variables y_j. It is easy to go from there to the *multinormal distribution,* where $\mathbf{y}$ is a p-dimensional random variable whose p dimensions are *not*

Multi-normal

independent. In order to do so, one simply replaces the above matrix $\boldsymbol{\Sigma}$ by a dispersion matrix containing variances and non-zero covariances, i.e. (eq. 4.2):

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

Using this dispersion matrix $\boldsymbol{\Sigma}$, eq. 4.19 now describes the probability density $f(\mathbf{y})$ for a p-dimensional multinormal distribution.

Given eq. 4.11, eq. 4.19 for point $\mathbf{y}$ may be rewritten as:

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{p/2}|\mathbf{D}(\sigma)||\mathbf{P}|^{1/2}} \exp\left\{-(1/2)\left[y-\mu\right]\mathbf{D}(\sigma)^{-1}\,\mathbf{P}^{-1}\,\mathbf{D}(\sigma)^{-1}\left[y-\mu\right]'\right\} \quad \textbf{(4.20)}$$

Replacing, in eq. 4.20, vector $\mathbf{y}$ from the p-dimensional matrix $\mathbf{Y}$ by vector $\mathbf{z}$ from the p-dimensional standardized matrix $\mathbf{Z}$ (eq. 1.12) gives:

$$f(\mathbf{z}) = \frac{1}{(2\pi)^{p/2}|\mathbf{P}|^{1/2}} \exp\left\{-(1/2)\,\mathbf{z}\,\mathbf{P}^{-1}\,\mathbf{z}'\right\} \quad \textbf{(4.21)}$$

because $\left[y-\mu\right]\mathbf{D}(\sigma)^{-1} = \mathbf{z}$ and, for a standardized variable $\mathbf{z}$, $\mathbf{D}(\sigma) = \mathbf{I}$.

Equation 4.21 stresses a fundamental point, which was already clear in eq. 4.20: *the correlations* ρ *are parameters of the multinormal distribution,* together with the means μ and standard deviations σ. This property of ρ is shown in Table 4.5.

Three sets of parameters are therefore necessary to specify a multidimensional normal distribution, i.e. the vector of *means* $\boldsymbol{\mu}$, the diagonal matrix of *standard deviations* $\mathbf{D}(\sigma)$, and the *correlation matrix* $\mathbf{P}$. In the unidimensional normal distribution (eq. 4.15), μ and σ were the only parameters because there is no correlation ρ for a single variable.

Bivariate normal

It is not possible to represent, in a plane, more than three dimensions. Thus, for the purpose of illustration, only the simplest case of multinormal distribution will be considered, i.e. the *bivariate normal distribution*, where:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 & \mu_2 \end{bmatrix} \qquad \mathbf{D}(\sigma) = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \qquad \mathbf{P} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

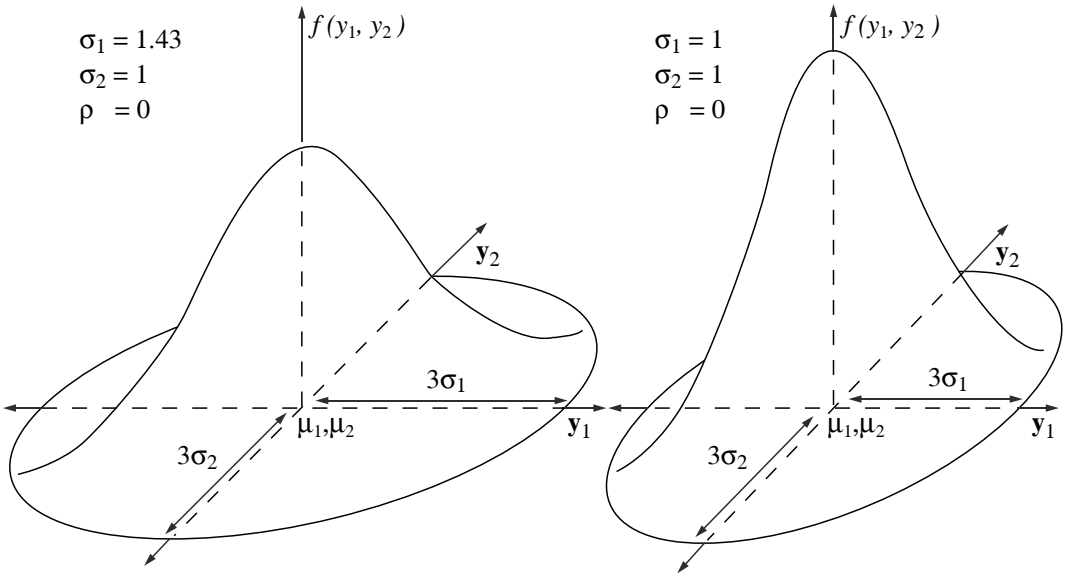


Figure 4.6 Roles of σ_1 and σ_2 in the bivariate normal distribution.

Since $|\mathbf{D}(\sigma)| = \sigma_1\sigma_2$ and $|\mathbf{P}| = (1 - \rho^2)$ in this case, eq. 4.20 becomes:

$$f(y_1,y_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\{-(1/2)\,[y-\mu]\,\mathbf{D}(1/\sigma)\,(1-\rho^2)^{-1}\begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}\,\mathbf{D}(1/\sigma)\,[y-\mu]'\}$$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2}\frac{1}{(1-\rho^2)}\left[\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{y_1-\mu_1}{\sigma_1}\right)\left(\frac{y_2-\mu_2}{\sigma_2}\right) + \left(\frac{y_2-\mu_2}{\sigma_2}\right)^2\right]\right\}$$

Figure 4.6 shows bivariate normal distributions with their typical "bell" shapes. The two examples illustrate the roles of σ_1 and σ_2. Further examination of the multinormal mathematics is required to specify the role of ρ.

Coming back to the probability density of the multidimensional distribution and neglecting the constant –1/2, the remainder of the exponent in eq. 4.19 is:

$$[y-\mu]\,\mathbf{\Sigma}^{-1}\,[y-\mu]'$$

When it is made equal to a positive constant (α), this algebraic form specifies the equation of any of the points $[y-\mu]$ on a *p*-dimensional *ellipse*:

$$[y-\mu]\,\mathbf{\Sigma}^{-1}\,[y-\mu]' = \alpha \qquad \textbf{(4.22)}$$

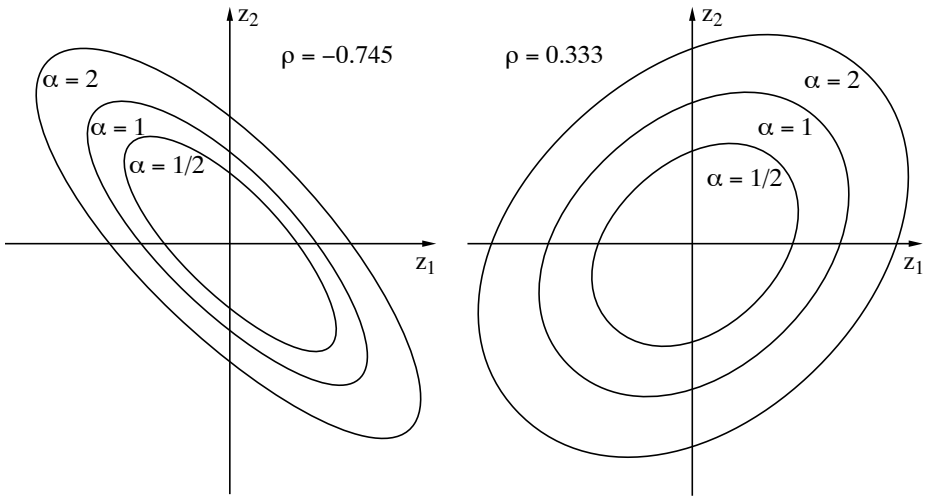


Figure 4.7 Concentration ellipses of a standardized bivariate normal distribution. Role of the correlation ρ.

A family of such multidimensional ellipses may be generated by varying the constant α. All these ellipses have the multidimensional point $\boldsymbol{\mu}$ as their common centre.

It is easy to understand the meaning of eq. 4.22 by examining the two-dimensional case. Without loss of generality, it is convenient to use the standardized variable (z_1,z_2) instead of (y_1,y_2). In that case, the family of *ellipses* (i.e. two-dimensional ellipsoids) is centred on the origin $\boldsymbol{\mu} = [0\ 0]$. For each point with coordinates $[z_1\ z_2]$, the exponent of the *standardized bivariate normal density* is (from expression on the previous page):

$$\frac{1}{1-\rho^2}[z_1^2 - 2\rho z_1 z_2 + z_2^2]$$

This exponent specifies, in two-dimensional space, the equation of a family of ellipses:

$$\frac{1}{1-\rho^2}[z_1^2 - 2\rho z_1 z_2 + z_2^2] = \alpha$$

$$z_1^2 - 2\rho z_1 z_2 + z_2^2 = \alpha(1-\rho^2)$$

Figure 4.7 illustrates the role played by ρ in determining the general shape of the family of ellipses. As ρ approaches zero, the ellipses tend to become circular. In contrast, as ρ approaches +1 or –1, the ellipses tend to elongate. The sign of ρ determines the orientation of the ellipses relative to the axes.

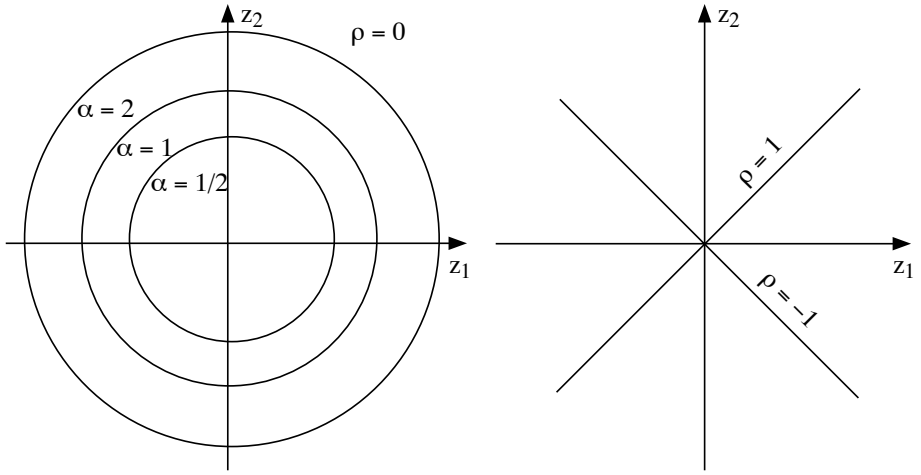


Figure 4.8 Concentration ellipses of a standardized bivariate normal distribution. Extreme values of the correlation ρ.

Actually, when $\rho = 0$ (Fig. 4.8, left), the equation for the family of ellipses becomes:

$$z_1^2 - [2 \times 0 \times z_1 z_2] + z_2^2 = \alpha(1 - 0)$$

or $z_1^2 + z_2^2 = \alpha$, which is the equation of a *circle*.

In contrast, when $\rho = \pm 1$, the equation becomes:

$$z_1^2 - [2 \times (\pm 1) \times z_1 z_2] + z_2^2 = \alpha[1 - (\pm 1)^2]$$

$$z_1^2 \mp 2 z_1 z_2 + z_2^2 = 0$$

hence $[z_1 \mp z_2]^2 = 0$, so that $z_1 \mp z_2 = 0$, and thus $z_1 = \pm z_2$,

which is the equation of a *straight line* with a positive or negative slope of 1 (±45° angle).

Such a family of ellipses, called *concentration ellipses*, is comparable to a series of contour lines on the two-dimensional normal distribution (Fig. 4.6). Increasing the value of α corresponds to moving down along the sides of the distribution. The concentration ellipses pass through points of equal probabilities around the bivariate normal distribution. The role of ρ then becomes clear: when $\rho = 0$, the "bell" of probability densities is perfectly circular (in overhead view); as ρ increases in absolute

value, the "bell" of the probability densities flattens out, until it becomes unidimensional when $\rho = \pm 1$. Indeed, when there is a perfect correlation between two dimensions (i.e. $\rho = \pm 1$), a single dimension, at an angle of 45° with respect to the two original variables, is sufficient to specify the joint distribution of probability densities.

When the number of dimensions is $p = 3$, the family of concentration ellipses becomes a family of concentration *ellipsoids* and, when $p > 3$, a family of *hyperellipsoids*. The meaning of these ellipsoids and hyperellipsoids is the same as in the two-dimensional case although it is not possible to draw them on a sheet of paper.

4.4 Principal axes

Various aspects of the multinormal distribution have been examined in the previous section. One of these, namely the *concentration ellipses* (Fig. 4.7), is the gateway to a topic of great importance for ecologists. In the present section, a method will be developed for determining the *principal axes* of the concentration hyperellipsoids; for simplicity, the term ellipsoid will be used in the following discussion. The *first principal axis* is the line that passes through the dimension of greatest variance of the ellipsoid. The next *principal axes* go through the next dimensions of greatest variance, smaller and smaller, of the p-dimensional ellipsoid. Hence, p consecutive principal axes are determined. These principal axes will be used, in Section 9.1, as the basis for *principal component analysis*.

In the two-dimensional case, the *first principal axis* corresponds to the *major axis* of the concentration ellipse and the *second principal axis* to the *minor axis*. These two axes are perpendicular to each other. Similarly in the p-dimensional case, there are p consecutive axes, which are all perpendicular to one another in the hyperspace.

The first principal axis goes through the p-dimensional centre $\boldsymbol{\mu} = [\mu_1\ \mu_2\ \ldots\ \mu_p]$ of the ellipsoid, and it crosses the surface of the ellipsoid at a point designated here by $\mathbf{y} = [y_1\ y_2\ \ldots\ y_p]$. The values of $\boldsymbol{\mu}$ and $\mathbf{y}$ specify a vector in the p-dimensional space (Section 2.4). The length of the axis, from $\boldsymbol{\mu}$ to the surface of the ellipsoid, is calculated using Pythagoras' formula:

$$[(y_1 - \mu_1)^2 + (y_2 - \mu_2)^2 + \ldots + (y_p - \mu_p)^2]^{1/2} = ([y - \mu][y - \mu]')^{1/2}$$

Actually, this is only half the length of the axis, which extends equally on both sides of $\boldsymbol{\mu}$. The coordinates of the *first* principal axis must be chosen in such a way as to maximize the length of the axis. This can be achieved by maximizing the square of the half-length:

$$[y - \mu][y - \mu]'$$

Calculating coordinates corresponding to the axis with the greatest length is subjected to the constraint that the end point **y** be on the surface of the ellipsoid. This constraint is made explicit using eq. 4.22, which specifies the ellipsoid:

$$[y - \mu]\ \boldsymbol{\Sigma}^{-1}\ [y - \mu]' = \alpha$$

$$[y - \mu]\ \boldsymbol{\Sigma}^{-1}\ [y - \mu]' - \alpha = 0$$

Principal axis

Lagrangian multipliers are used to compute the maximum and minimum values of a function of several variables when the relationships among the variables are known. In the present case, the above two equations, which describe the square of the half-length of the first principal axis and the constraint, are combined into a single function:

$$f(\mathbf{y}) = [y - \mu][y - \mu]' - \lambda\ \{[y - \mu]\ \boldsymbol{\Sigma}^{-1}\ [y - \mu]' - \alpha\}$$

Lagrangian multiplier

Scalar λ is called a *Lagrangian multiplier*[*]. The values that maximize this function are found by the usual method of setting the equation's partial derivative equal to 0:

$$\frac{\partial}{\partial \mathbf{y}} f(\mathbf{y}) = \mathbf{0}$$

$$\frac{\partial}{\partial \mathbf{y}} [y - \mu][y - \mu]' - \lambda\ \frac{\partial}{\partial \mathbf{y}} \{[y - \mu]\ \boldsymbol{\Sigma}^{-1}\ [y - \mu]' - \alpha\} = \mathbf{0}$$

It is important to remember here that **y** is a *p*-dimensional *vector* $(y_1, y_2, \ldots, y_p)$, which means that the above equation is successively derived with respect to y_1, y_2, … and y_p. Therefore, derivation with respect to **y** represents in fact calculating a series of *p* partial derivatives (∂y_j). The results of the derivation may be rewritten as a (column) vector with *p* elements:

$$2\ [y - \mu] - 2\ \lambda\ \boldsymbol{\Sigma}^{-1}\ [y - \mu] = \mathbf{0}$$

One may factor out $[y - \mu]$ and eliminate the constant 2:

$$(\mathbf{I} - \lambda\ \boldsymbol{\Sigma}^{-1})\ [y - \mu] = \mathbf{0}$$

Multiplying both sides of the equation by $\boldsymbol{\Sigma}$ gives:

$$(\boldsymbol{\Sigma} - \lambda \mathbf{I})\ [y - \mu] = \mathbf{0} \qquad \textbf{(4.23)}$$

The general equation defining eigenvectors (eq. 2.22) is $(\mathbf{A} - \lambda \mathbf{I})\ \mathbf{u} = \mathbf{0}$. Replacing, in that equation, **A** by $\boldsymbol{\Sigma}$ and **u** by $[y - \mu]$ produces eq. 4.23. This leads to the conclusion that the *vector of coordinates that specifies the first principal axis is one of the eigenvectors* $[y - \mu]$ *of matrix* $\boldsymbol{\Sigma}$.

[*] After Joseph-Louis Lagrange (1736-1813), mathematician and astronomer.

In order to find out which of the p eigenvectors of $\mathbf{\Sigma}$ is the vector of coordinates of the *first principal axis,* go back to the equation resulting from the partial derivation (above) and transfer the second term to the right, after eliminating the constant 2:

$$[y - \mu] = \lambda\, \mathbf{\Sigma}^{-1}\, [y - \mu]$$

The two sides are then premultiplied by $[y - \mu]'$:

$$[y - \mu]'\, [y - \mu] = \lambda\, [y - \mu]'\, \mathbf{\Sigma}^{-1}\, [y - \mu]$$

Since $[y - \mu]'\, \mathbf{\Sigma}^{-1}\, [y - \mu] = \alpha$ (eq. 4.22), it follows that:

$$[y - \mu]'\, [y - \mu] = \lambda\alpha$$

Considering the first eigenvalue λ_1, the term on the left-hand side of the equation is the square of the half-length of the first principal axis (see above). Thus, for a given value α, the length of the *first principal axis* is maximized by taking the largest possible value for λ or, in other words, the *largest eigenvalue,* λ_1, *of matrix* $\mathbf{\Sigma}$. The vector of coordinates of the *first principal axis* is therefore the eigenvector corresponding to the largest eigenvalue of $\mathbf{\Sigma}$.

Eigenvalue

Numerical example. The above equations are illustrated using the bivariate data matrix from Section 9.1 (principal component analysis). The sample covariance matrix is:

$$\mathbf{S} = \begin{bmatrix} 8.2 & 1.6 \\ 1.6 & 5.8 \end{bmatrix}$$

There are two eigenvalues, $\lambda_1 = 9$ and $\lambda_2 = 5$, computed using eq. 2.23. To normalize the eigenvectors (written as column vectors), put $[y - \mu]'\, [y - \mu] = \lambda\alpha = 1$ for each of them; in other words, $\alpha_1 = 1/9$ and $\alpha_2 = 1/5$. The normalized eigenvectors were called $\mathbf{y}_1$ and $\mathbf{y}_2$ until now in this section; they will be denoted $\mathbf{u}_j$ from now on, as in Sections 2.9 and 2.10. They form matrix $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2]$:

$$\mathbf{y}_1 = \mathbf{u}_1 = \begin{bmatrix} 0.8944 \\ 0.4472 \end{bmatrix} \text{ and } \mathbf{y}_2 = \mathbf{u}_2 = \begin{bmatrix} -0.4472 \\ 0.8944 \end{bmatrix}$$

These eigenvectors are of length 1 since they have been normalized. They determine the *directions* of the major and minor axes of the bivariate distribution. The matrix of eigenvectors $\mathbf{U}$ must be multiplied by the diagonal matrix containing the square roots of the eigenvalues ($\mathbf{U\Lambda}^{1/2}$, eq. 9.10) to provide a new matrix whose columns give the coordinates where the two principal axes cross an ellipsoid with size $\alpha = 1$. This example is further developed in Chapter 9.

To find the vectors of coordinates specifying the p successive principal axes,

• rank the p eigenvalues of matrix $\mathbf{\Sigma}$ in decreasing order:

$$\lambda_1 > \lambda_2 > \ldots > \lambda_p \geq 0$$

Note that the eigenvalues of a matrix $\mathbf{\Sigma}$ are all positive (end of Section 4.1);

• associate the p eigenvectors to their corresponding eigenvalues. The orientation of the p successive principal axes is given by the eigenvectors, which are associated with the p eigenvalues ranked in decreasing order. The eigenvectors of a covariance matrix $\mathbf{\Sigma}$ are orthogonal to one another because $\mathbf{\Sigma}$ is symmetric (Section 2.9). In the case of multiplicity (Section 2.10, Third property), the orthogonal axes may be rotated to an infinity of "principal" directions, i.e. two equal λ's result in a circle and several determine a hypersphere (multidimensional sphere) where no orientation prevails.

The next step consists in calculating a new p-dimensional set of variables, forming matrix $\mathbf{V}$, that position the dispersion ellipses with respect to the principal axes instead of the original Cartesian system. $\mathbf{V}$ is related to the original data matrix $\mathbf{Y}$ (eq. 4.1) through the following transformation:

$$\mathbf{V} = [y - \mu]\,\mathbf{U} \qquad \textbf{(4.24)}$$

where each of the p columns in matrix $\mathbf{U}$ is the normalized eigenvector $\mathbf{u}_k$ corresponding to the k-th principal axis. Because vectors $\mathbf{u}_k$ are both orthogonal and normalized, matrix $\mathbf{U}$ is said to be *orthonormal* (Section 2.8). This transformation results in shifting the origin of the system of axes to the p-dimensional point $\boldsymbol{\mu}$ followed by a rigid rotation of the translated axes into the principal axes (Fig. 4.9), which form matrix $\mathbf{V}$.

The dispersion matrix of $\mathbf{V}$ is:

$$\mathbf{\Sigma}_\mathbf{V} = \frac{1}{(n-1)}\,(\mathbf{V}'\mathbf{V}) \;=\; \frac{1}{(n-1)}\mathbf{U}'\;[y-\mu]'\,[y-\mu]\,\mathbf{U} = \mathbf{U}'\mathbf{\Sigma}\mathbf{U}$$

where $\mathbf{\Sigma}$ is the dispersion matrix of the original matrix $\mathbf{Y}$. So, the *variance* of the k-th dimension $\mathbf{v}_k$ (i.e. the k-th principal axis) is:

$$s^2(\mathbf{v}_k) = \mathbf{u}_k'\,\mathbf{\Sigma}\mathbf{u}_k$$

Since, by definition, $\mathbf{\Sigma}\mathbf{u}_k = \lambda_k\mathbf{u}_k$ (eq. 2.21) and $\mathbf{u}_k'\,\mathbf{u}_k = 1$, it follows that:

$$s^2(\mathbf{v}_k) = \mathbf{u}_k'\,\mathbf{\Sigma}\mathbf{u}_k = \mathbf{u}_k'\,\lambda_k\mathbf{u}_k = \lambda_k\;\mathbf{u}_k'\,\mathbf{u}_k = \lambda_k\,(1) = \lambda_k \qquad \textbf{(4.25)}$$

with $\lambda_k \geq 0$ in all cases since $\mathbf{\Sigma}$ is positive semi-definite. The *covariance* of any two vectors of matrix $\mathbf{V}$ is zero because the product of two orthogonal vectors $\mathbf{u}_k$ and $\mathbf{u}_h$ is zero (Section 2.5):

$$s(\mathbf{v}_k,\mathbf{v}_h) = \mathbf{u}_k'\,\mathbf{\Sigma}\mathbf{u}_h = \mathbf{u}_k'\,\lambda_h\mathbf{u}_h = \lambda_h\;\mathbf{u}_k'\,\mathbf{u}_h = \lambda_k\,(0) = 0 \qquad \textbf{(4.26)}$$

The last two points are of utmost importance, since they are the basis for using the principal axes (and thus principal component analysis; Section 9.1) in ecology: (1) *the variance of a principal axis is equal to the eigenvalue associated with that axis*

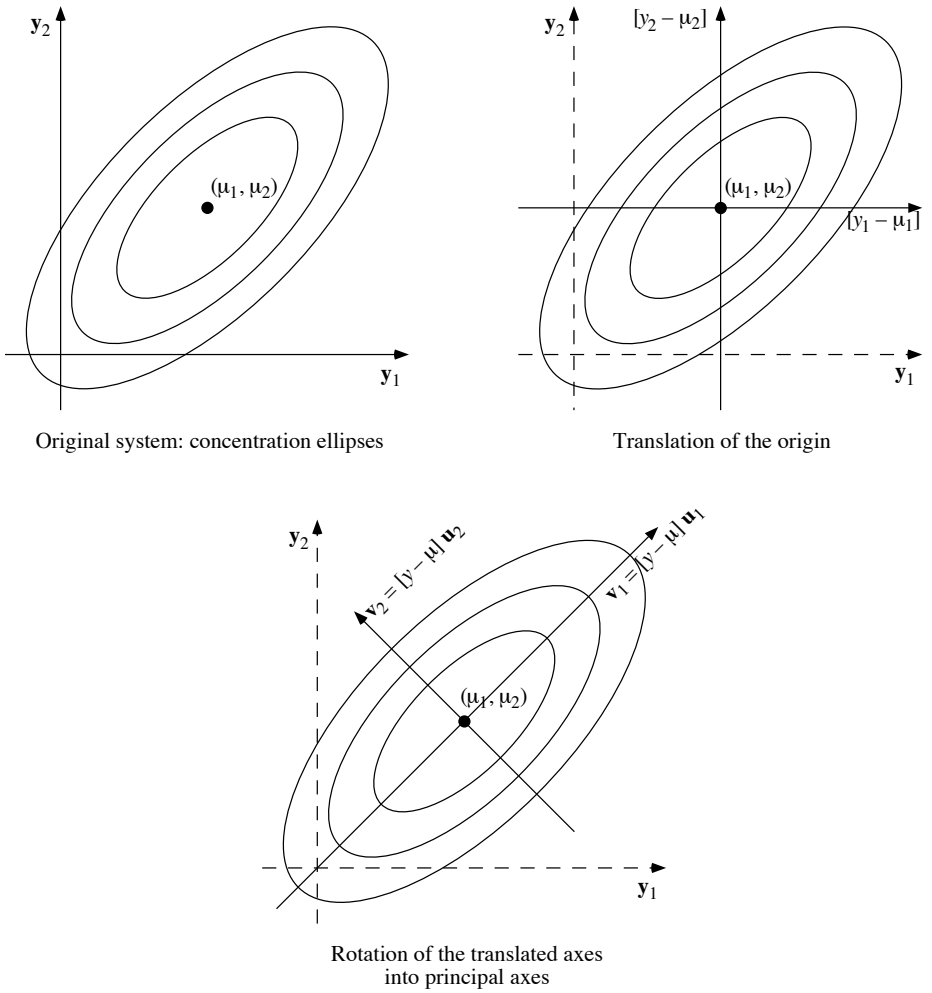


Figure 4.9 Result of the transformation $\mathbf{V} = [y - \mu]\,\mathbf{U}$ (eq. 4.24)**.**

(eq. 4.25) and (2) *the p dimensions of the transformed variable are linearly independent* since their covariances are zero (eq. 4.26).

A last point concerns the meaning of the p elements u_{jk} of the normalized eigenvectors $\mathbf{u}_k$. The values of these elements determine the rotation of the system of axes, so that they correspond to angles. Figure 4.10 illustrates, for the two-dimensional case, how the elements of the eigenvectors are related to the rotation angles. Using the

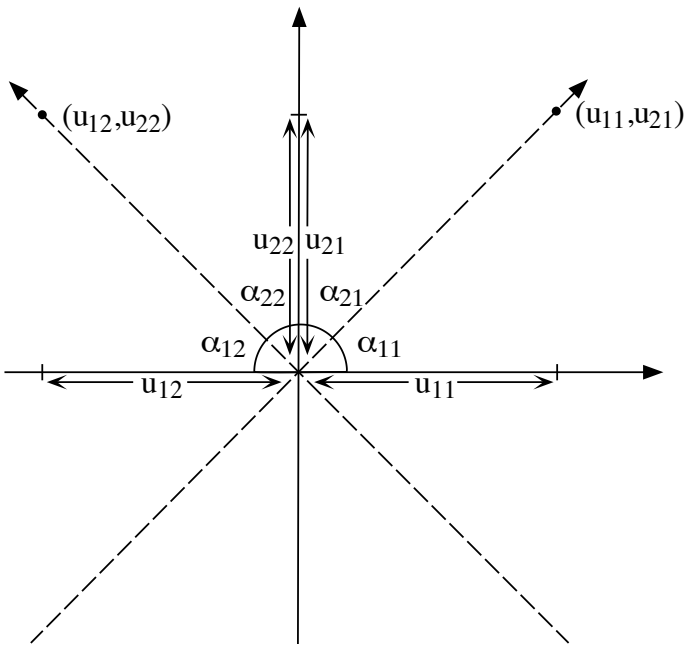


Figure 4.10 Geometrical meaning of the principal axes.

trigonometric functions for right-angled triangles, the angular relationships in Fig. 4.10 may be rewritten as cosines:

$$\cos \alpha_{11} = \text{length } u_{11} \text{ / length of vector } (u_{11}, u_{21}) = u_{11}$$

$$\cos \alpha_{21} = \text{length } u_{21} \text{ / length of vector } (u_{11}, u_{21}) = u_{21}$$

$$\cos \alpha_{12} = \text{length } u_{12} \text{ / length of vector } (u_{12}, u_{22}) = u_{12}$$

$$\cos \alpha_{22} = \text{length } u_{22} \text{ / length of vector } (u_{12}, u_{22}) = u_{22}$$

because the lengths of the *normalized* vectors (u_{11}, u_{21}) and (u_{12}, u_{22}) are 1 (Section 2.4). Eigenvector $\mathbf{u}_k$ determines the direction of the k-th main axis; it follows from the above trigonometric relationships that elements u_{jk} of the normalized eigenvectors are *direction cosines*. Each direction cosine specifies the angle between an original Cartesian axis j and a principal axis k.

Direction cosine

The two-dimensional case, illustrated in Figs. 4.9 and 4.10, is the simplest to compute. The standardized dispersion matrix is of the general form:

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

When ρ is positive, the eigenvalues of $\mathbf{P}$ are $\lambda_1 = (1 + \rho)$ and $\lambda_2 = (1 - \rho)$. The *normalized* eigenvectors are:

$$\mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \qquad \mathbf{u}_2 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

Therefore, the first principal axis goes through the point $(1/\sqrt{2}\ ,\ 1/\sqrt{2}\ )$, so that it cuts the first and third quadrants at a 45° angle. Its direction cosines are $\cos \alpha_{11} = 1/\sqrt{2}$ and $\cos \alpha_{12} = 1/\sqrt{2}$, which indeed specify 45° angles with respect to the two axes of the first quadrant. The second principal axis goes through $(-1/\sqrt{2}\ ,\ 1/\sqrt{2}\ )$, so that it cuts the second and fourth quadrants at 45°. Its direction cosines are $\cos \alpha_{21} = -1/\sqrt{2}$ and $\cos \alpha_{22} = 1/\sqrt{2}$, which determine 45°angles with respect to the two axes of the second quadrant.

When ρ is *negative,* the eigenvalues of $\mathbf{P}$ are $\lambda_1 = (1 - \rho)$ and $\lambda_2 = (1 + \rho)$. Consequently the first principal axis goes through $(-1/\sqrt{2}\ ,\ 1/\sqrt{2}\ )$ in the second quadrant, while the second principal axis with coordinates $(1/\sqrt{2}\ ,\ 1/\sqrt{2}\ )$ cuts the first quadrant. A value $\rho = 0$ entails a case of multiplicity since $\lambda_1 = \lambda_2 = 1$. This results in an infinite number of "principal" axes, i.e. any two perpendicular diameters would fit the circular concentration ellipse (Fig. 4.8, left).

These concepts, so far quite abstract, will find direct applications to ecology in Section 9.1, where principal component analysis is described.

4.5 Multiple and partial correlations

Section 4.2 considered, in a multidimensional context, the correlation of pairs of variables, which represent two dimensions of a *p*-dimensional random variable. However, the multidimensional nature of ecological data allows other approaches to correlation analysis. These statistics are examined in the present section and compared graphically in Box 4.1

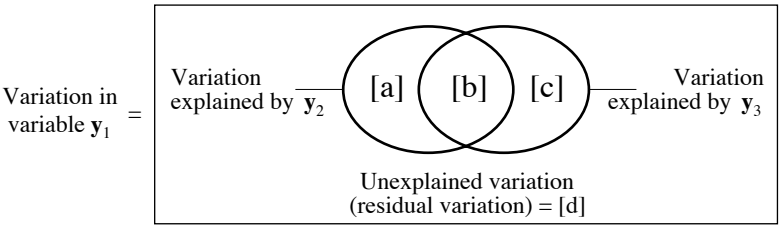
The following developments will require that the *p*-dimensional correlation matrix $\mathbf{R}$ be partitioned into four submatrices. Indices assigned to the submatrices follow the general convention on matrix indices (Section 2.1):

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{bmatrix} \qquad \textbf{(4.27)}$$

Variation partitioning — Box 4.1

Variation partitioning, which is described in detail in Subsection 10.3.5, provides a general framework to illustrate the similarities and differences between the coefficient of multiple determination and the partial and semipartial correlation coefficients, as well as the corresponding F-statistics.

Three variables only, $\mathbf{y}_1$, $\mathbf{y}_2$, and $\mathbf{y}_3$, are considered in this example. In the following Venn diagram, the rectangle represents the total sum of squares of variable $\mathbf{y}_1$:



In the multiple regression of $\mathbf{y}_1$ on $\mathbf{y}_2$ and $\mathbf{y}_3$, $\hat{y}_1 = b_0 + b_2 y_2 + b_3 y_3$ (this is an application of eq. 10.15), the coefficient of multiple determination, which is the square of the coefficient of multiple correlation, is:

$$R^2_{1.23} = \frac{[a + b + c]}{[a + b + c + d]} \qquad \text{with} \quad F = \frac{[a + b + c]/2}{[d]/(n-3)}$$

The partial correlation of $\mathbf{y}_1$ with $\mathbf{y}_2$ while controlling for the effect of $\mathbf{y}_3$ is:

$$r_{12.3} = \sqrt{\frac{[a]}{[a + d]}} \qquad \text{with} \quad F = \frac{[a]/1}{[d]/(n-3)}$$

The semipartial correlation of $\mathbf{y}_1$ with $\mathbf{y}_2$ in the presence of $\mathbf{y}_3$ is:

$$r_{1(2.3)} = \sqrt{\frac{[a]}{[a+b+c+d]}} \qquad \text{with} \quad F = \frac{[a]/1}{[d]/(n-3)}$$

The coefficients of partial and semipartial correlation receive the same sign as the corresponding coefficient of partial regression.

The test of a partial regression coefficient, b_2 or b_3, is the same (i.e. it has the same F-statistic) as the test of the corresponding partial correlation coefficient, $r_{12.3}$ or $r_{13.2}$. The F-statistic is always the ratio of two *independent* portions of the variation of $\mathbf{y}_1$, each one divided by its degrees of freedom; see eqs. 4.39 and 4.40.

There are two possible approaches to linear correlation involving several variables or several dimensions of a multidimensional variable. The first one, which is called *multiple* (*linear*) *correlation*, measures the intensity of the relationship between a *response* variable and a linear combination of several *explanatory* variables. The second approach, called *partial* (*linear*) *correlation*, measures the intensity of the linear relationship between two variables, while taking into account their relationships with other variables.

1 — Multiple linear correlation

Multiple correlation applies to cases where there is one *response* variable and several *explanatory* variables. This situation is further studied in Section 10.3, within the context of *multiple regression*. The *coefficient of multiple determination* (R^2; eq. 10.20) measures the fraction of the variance of $\mathbf{y}_k$ that is explained by a linear combination of $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_j, \ldots$ and $\mathbf{y}_p$:

$$R^2_{k.12..j..p} = \frac{b_1 s_{1k} + b_2 s_{2k} + \ldots + b_j s_{jk} + \ldots + b_p s_{pk}}{s^2_k} \qquad \textbf{(4.28)}$$

where p is here the number of explanatory variables.The concept is illustrated in Box 4.1. In eq. 4.28, coefficients b are the coefficients of the multiple regression (Subsection 10.3.3) of $\mathbf{y}_k$ on the explanatory variables. A coefficient $R^2_{k.12..j..p} = 0.73$, for example, means that the linear relationships of variables $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_j, \ldots$ and $\mathbf{y}_p$ with $\mathbf{y}_k$ explain 73% of the variation of $\mathbf{y}_k$ around its mean. The *multiple correlation coefficient* (R) is the square root of the coefficient of multiple determination:

$$R_{k.12..j..p} = \sqrt{R^2_{k.12..j..p}} \qquad \textbf{(4.29)}$$

To calculate R^2 using matrix algebra, a correlation matrix $\mathbf{R}$ is written for variables $\mathbf{y}_k$ and $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_j, \ldots, \mathbf{y}_p\}$, with $\mathbf{y}_k$ in the first position. Partitioning this matrix following eq. 4.27 to compute a multiple correlation coefficient gives:

$$\mathbf{R} = \left[\begin{array}{c|ccc} 1 & r_{k1} & r_{k2} \ldots & r_{kp} \\ \hline r_{1k} & 1 & r_{12} \ldots & r_{1p} \\ r_{2k} & r_{21} & 1 \ldots & r_{2p} \\ \cdot & \cdot & \cdot \ldots & \cdot \\ \cdot & \cdot & \cdot \ldots & \cdot \\ \cdot & \cdot & \cdot \ldots & \cdot \\ r_{pk} & r_{p1} & r_{p2} \ldots & 1 \end{array}\right] = \begin{bmatrix} 1 & \mathbf{r}_{12} \\ \mathbf{r}_{21} & \mathbf{R}_{22} \end{bmatrix} \qquad \textbf{(4.30)}$$

where $\mathbf{r}_{12} = \mathbf{r}'_{21}$ is a vector containing the correlation coefficients $r_{k1}, r_{k2}, \ldots, r_{kp}$. Using $\mathbf{r}_{12}, \mathbf{r}_{21}$ and $\mathbf{R}_{22}$ as defined in eq. 4.30, R^2 is calculated as:

$$R^2 = \mathbf{r}_{12}\mathbf{R}_{22}^{-1}\mathbf{r}_{21} = \mathbf{r}'_{21}\mathbf{R}_{22}^{-1}\mathbf{r}_{21} \qquad \textbf{(4.31)}$$

Equation 4.31 is expanded using eq. 2.17:

$$R^2 = \mathbf{r}'_{21}\mathbf{R}_{22}^{-1}\mathbf{r}_{21} = \mathbf{r}'_{21}\frac{1}{|\mathbf{R}_{22}|}\begin{bmatrix} \text{cof}(r_{11}) & \text{cof}(r_{21}) & \ldots & \text{cof}(r_{p1}) \\ \text{cof}(r_{12}) & \text{cof}(r_{22}) & \ldots & \text{cof}(r_{p2}) \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \text{cof}(r_{1p}) & \text{cof}(r_{2p}) & \ldots & \text{cof}(r_{pp}) \end{bmatrix}\mathbf{r}_{21}$$

$$R^2 = \frac{1}{|\mathbf{R}_{22}|}\left(|\mathbf{R}_{22}| - |\mathbf{R}|\right) = 1 - \frac{|\mathbf{R}|}{|\mathbf{R}_{22}|} \qquad \textbf{(4.32)}$$

As an exercise, it is easy to check that

$$|\mathbf{R}_{22}| - |\mathbf{R}| = \mathbf{r}'_{21}\ [\text{adjugate matrix of } \mathbf{R}_{22}]\ \mathbf{r}_{21}$$

Multiple correlation

The *coefficient of multiple correlation* is calculated from eqs. 4.31 or 4.32:

$$R_{k.12\ldots j\ldots p} = \sqrt{\mathbf{r}'_{21}\mathbf{R}_{22}^{-1}\mathbf{r}_{21}} \quad \text{or} \quad R_{k.12\ldots j\ldots p} = \sqrt{1 - \frac{|\mathbf{R}|}{|\mathbf{R}_{22}|}} \qquad \textbf{(4.33)}$$

A third way of calculating R^2 is described in eq. 4.38, near the end of Subsection 4.5.2 on partial correlation.

When two or more variables in matrix $\mathbf{R}_{22}$ are perfectly correlated (i.e. $r = 1$ or $r = -1$), the rank of $\mathbf{R}_{22}$ is smaller than its order (Section 2.7), hence $|\mathbf{R}_{22}| = 0$. Calculation of R thus requires the elimination of redundant variables from matrix $\mathbf{R}$.

Numerical example. A simple example, with three variables ($\mathbf{y}_1$, $\mathbf{y}_2$ and $\mathbf{y}_3$), illustrates the above equations. Matrix $\mathbf{R}$ is:

$$\mathbf{R} = \begin{bmatrix} 1 & 0.4 & 0.8 \\ 0.4 & 1 & 0.5 \\ 0.8 & 0.5 & 1 \end{bmatrix}$$

The *coefficient of multiple determination* $R^2_{1.23}$ is first calculated using eq. 4.31:

$$R^2_{1.23} = \begin{bmatrix} 0.4 & 0.8 \end{bmatrix} \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0.4 \\ 0.8 \end{bmatrix}$$

$$R^2_{1.23} = \begin{bmatrix} 0.4 & 0.8 \end{bmatrix} \begin{bmatrix} 1.33 & -0.67 \\ -0.67 & 1.33 \end{bmatrix} \begin{bmatrix} 0.4 \\ 0.8 \end{bmatrix}$$

$$R^2_{1.23} = 0.64$$

Equation 4.32 leads to an identical result:

$$R^2_{1.23} = 1 - \frac{\begin{vmatrix} 1 & 0.4 & 0.8 \\ 0.4 & 1 & 0.5 \\ 0.8 & 0.5 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & 0.5 \\ 0.5 & 1 \end{vmatrix}}$$

$$R^2_{1.23} = 1 - \frac{0.27}{0.75} = 0.64$$

The linear combination of variables $\mathbf{y}_2$ and $\mathbf{y}_3$ explains 64% of the variance of $\mathbf{y}_1$. The *multiple correlation coefficient* is $R_{1.23} = 0.8$.

2 — Partial correlation

The second approach to correlation, in the multidimensional context, applies to situations where the relationship between two variables is influenced by their relationships with other variables. Two coefficients are described in Box 4.1: the *partial* and *semipartial correlation coefficients*.

The *partial correlation coefficient* is related to partial multiple regression (Subsection 10.3.5). It measures what the correlation between y_j and y_k would be if other variables y_1, y_2, … and y_p, hypothesized to influence both y_j and y_k, were held constant at their means. The partial correlation between variables y_j and y_k, when controlling for their relationships with y_1, y_2, … and y_p, is written $r_{jk.12\ldots p}$.

In order to calculate the partial correlation coefficients, the set of variables is divided into two subsets. The *first subset* contains the variables between which the partial correlation is to be computed while controlling for the influence of the variables

in the second subset. The *second subset* thus contains the variables whose influence is to be taken into account. Matrix $\mathbf{R}$ is partitioned as follows (eq. 4.27):

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{bmatrix}$$

$\mathbf{R}_{11}$ (of order 2 × 2 for partial correlations) and $\mathbf{R}_{22}$ contain the correlations among variables in the first and the second subsets, respectively, whereas $\mathbf{R}_{12}$ and $\mathbf{R}_{21}$ both contain the correlations between variables of the two subsets; $\mathbf{R}_{12} = \mathbf{R}'_{21}$.

The number of variables in the second subset determines the *order* of the partial correlation coefficient. This order is the number of variables whose effects are eliminated from the correlation between $\mathbf{y}_j$ and $\mathbf{y}_k$. For example $r_{12.345}$ (third-order partial correlation coefficient) means that the correlation between variables $\mathbf{y}_1$ and $\mathbf{y}_2$ is calculated while controlling for the linear effects of $\mathbf{y}_3$, $\mathbf{y}_4$, and $\mathbf{y}_5$.

The computation consists in subtracting from $\mathbf{R}_{11}$ (correlation matrix among the variables in the first subset) a second matrix containing the coefficients of multiple determination of the variables in the second subset on those in the first subset. These coefficients measure the fraction of the variance and covariance of the variables in the first subset that is explained by linear combinations of the variables in the second subset. They are computed by replacing in eq. 4.31 vector $\mathbf{r}_{21}$ by submatrix $\mathbf{R}_{21}$:

$$\mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{21} = \mathbf{R}'_{21}\mathbf{R}_{22}^{-1}\mathbf{R}_{21}$$

Subtracting this expression from $\mathbf{R}_{11}$ gives the *matrix of conditional correlations*:

$$\text{Matrix of conditional correlations} = \mathbf{R}_{11} - \mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{21} \qquad \textbf{(4.34)}$$

It can be shown that the maximum likelihood estimate ($\mathbf{R}_{1.2}$) of the partial correlation matrix $\mathbf{P}_{1.2}$ is:

$$\mathbf{R}_{1.2} = \mathbf{D}\,(r_{1.2})^{-1/2}\,(\mathbf{R}_{11} - \mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{21})\,\mathbf{D}\,(r_{1.2})^{-1/2} \qquad \textbf{(4.35)}$$

where $\mathbf{D}(r_{1.2})$ is the matrix of diagonal elements of the matrix of conditional correlation (eq. 4.34).

The calculation is illustrated for the three-dimensional case, in which there is a single controlled variable $\mathbf{y}_3$:

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{bmatrix}$$

This development will provide the algebraic formula for the partial correlation coefficients of order 1. Coefficients pertaining to variables of the first subset ($\mathbf{y}_1$ and $\mathbf{y}_2$) are in the first two rows and columns. Using eq. 4.35 gives:

$$\mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{21} = \begin{bmatrix} r_{13} \\ r_{23} \end{bmatrix} \begin{bmatrix} 1 \end{bmatrix}^{-1} \begin{bmatrix} r_{31} & r_{32} \end{bmatrix} = \begin{bmatrix} r_{13}^2 & r_{13}r_{23} \\ r_{13}r_{23} & r_{23}^2 \end{bmatrix}$$

$$\mathbf{R}_{11} - \mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{21} = \begin{bmatrix} 1 & r_{12} \\ r_{21} & 1 \end{bmatrix} - \begin{bmatrix} r_{13}^2 & r_{13}r_{23} \\ r_{13}r_{23} & r_{23}^2 \end{bmatrix} = \begin{bmatrix} (1-r_{13}^2) & (r_{12}-r_{13}r_{23}) \\ (r_{12}-r_{13}r_{23}) & (1-r_{23}^2) \end{bmatrix}$$

$$\mathbf{R}_{1.2} = \begin{bmatrix} 1/\sqrt{1-r_{13}^2} & 0 \\ 0 & 1/\sqrt{1-r_{23}^2} \end{bmatrix} \begin{bmatrix} (1-r_{13}^2) & (r_{12}-r_{13}r_{23}) \\ (r_{12}-r_{13}r_{23}) & (1-r_{23}^2) \end{bmatrix} \begin{bmatrix} 1/\sqrt{1-r_{13}^2} & 0 \\ 0 & 1/\sqrt{1-r_{23}^2} \end{bmatrix}$$

$$\mathbf{R}_{1.2} = \begin{bmatrix} 1 & \dfrac{r_{12}-r_{13}r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}} \\ \dfrac{r_{12}-r_{13}r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}} & 1 \end{bmatrix} = \begin{bmatrix} 1 & r_{12.3} \\ r_{12.3} & 1 \end{bmatrix}$$

The previous matrix equation provides the formula for the first-order partial correlation coefficient:

$$r_{12.3} = \frac{r_{12}-r_{13}r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}} \qquad \textbf{(4.36)}$$

The general formula, for coefficients of order p, is:

$$r_{jk.1\ldots p} = \frac{r_{jk.1\ldots(p-1)} - r_{jp.1\ldots(p-1)}r_{kp.1\ldots(p-1)}}{\sqrt{1-r_{jp.1\ldots(p-1)}^2}\sqrt{1-r_{kp.1\ldots(p-1)}^2}} \qquad \textbf{(4.37)}$$

When there are four variables, it is possible to calculate 12 first-order and 6 second-order partial correlation coefficients. Computing a second-order coefficient necessitates the calculation of 3 first-order coefficients. For example:

$$r_{12.34} = \frac{r_{12.3}-r_{14.3}r_{24.3}}{\sqrt{1-r_{14.3}^2}\sqrt{1-r_{24.3}^2}} = r_{12.43} = \frac{r_{12.4}-r_{13.4}r_{23.4}}{\sqrt{1-r_{13.4}^2}\sqrt{1-r_{23.4}^2}}$$

It is thus possible, as the number of variables increases, to calculate higher-order coefficients. Computing a coefficient of a given order requires the calculation of three

coefficients of the previous order, each of these requiring the calculation of coefficients of the previous order, and so on depending on the number of variables involved. Such a cascade of calculations is advantageously replaced by the direct matrix approach of eq. 4.35.

Numerical example. Partial correlations are calculated on the simple example already used for multiple correlation. Matrix **R** is:

$$\mathbf{R} = \begin{bmatrix} 1 & 0.4 & 0.8 \\ 0.4 & 1 & 0.5 \\ 0.8 & 0.5 & 1 \end{bmatrix}$$

Two subsets are formed, the first one containing descriptors $\mathbf{y}_1$ and $\mathbf{y}_2$ (between which the partial correlation is computed) and the second one $\mathbf{y}_3$ (whose influence on r_{12} is controlled for). Computations follow eqs. 4.34 and 4.35:

eq. 4.34 Matrix of conditional correlations $= \begin{bmatrix} 1 & 0.4 \\ 0.4 & 1 \end{bmatrix} - \begin{bmatrix} 0.8 \\ 0.5 \end{bmatrix} \begin{bmatrix} 1 \end{bmatrix}^{-1} \begin{bmatrix} 0.8 & 0.5 \end{bmatrix}$

$$= \begin{bmatrix} 1 & 0.4 \\ 0.4 & 1 \end{bmatrix} - \begin{bmatrix} 0.64 & 0.40 \\ 0.40 & 0.25 \end{bmatrix} = \begin{bmatrix} 0.36 & 0 \\ 0 & 0.75 \end{bmatrix}$$

eq. 4.35 $$\mathbf{R}_{1.2} = \begin{bmatrix} 1.67 & 0 \\ 0 & 1.15 \end{bmatrix} \begin{bmatrix} 0.36 & 0 \\ 0 & 0.75 \end{bmatrix} \begin{bmatrix} 1.67 & 0 \\ 0 & 1.15 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Thus, the partial correlation $r_{12.3} = 0$; this was unexpected given that $r_{12} = 0.4$. In other words, fraction [a] displayed in Box 4.1 is 0. The conclusion is that, when their (linear) relationships with $\mathbf{y}_3$ are taken into account, descriptors $\mathbf{y}_1$ and $\mathbf{y}_2$ are (linearly) independent. Similar calculations for the other two pairs of descriptors give: $r_{13.2} = 0.76$ and $r_{23.1} = 0.33$. The interpretation of these correlation coefficients will be further discussed in Subsection 4.5.4.

There is a relationship between the coefficients of *multiple* and *partial* correlation. The equation linking the two types of coefficients can be easily derived; in the multiple correlation equation, p is the number of variables other than $\mathbf{y}_k$:

when $p = 1$, the fraction of the variance of $\mathbf{y}_k$ not explained by $\mathbf{y}_1$ is the complement of the coefficient of determination $(1 - r_{k1}^2)$; this expression is sometimes called the *coefficient of nondetermination*;

Nondetermination

when $p = 2$, the fraction of the variance of $\mathbf{y}_k$ not explained by $\mathbf{y}_2$, while controlling for the linear influence of $\mathbf{y}_1$, is $(1 - r_{k2.1}^2)$, so that the fraction of the variance of $\mathbf{y}_k$ not explained by $\mathbf{y}_1$ and $\mathbf{y}_2$ is $(1 - r_{k1}^2)\,(1 - r_{k2.1}^2)$.

This leads to a general expression for the fraction of the variance of $\mathbf{y}_k$ that is not explained by $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_j, \ldots$ and $\mathbf{y}_p$:

$$(1 - r_{k1}^2)\ (1 - r_{k2.1}^2) \ldots (1 - r_{kj.12\ldots}^2) \ldots (1 - r_{kp.12\ldots j\ldots(p-1)}^2)$$

Multiple determination

The fraction of the variance of $\mathbf{y}_k$ that is explained by $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_j, \ldots$ and $\mathbf{y}_p$, i.e. the *coefficient of multiple determination* (square of the *multiple correlation*), is thus:

$$R_{k.12\ldots p}^2 = 1 - [\,(1 - r_{k1}^2)\ (1 - r_{k2.1}^2) \ldots (1 - r_{kp.12\ldots p-1}^2)\,] \qquad \textbf{(4.38)}$$

Numerical example. The same example as above is used to illustrate the calculation of the multiple correlation coefficient, using eq. 4.38:

$$R_{1.23}^2 = 1 - [\,(1 - r_{12}^2)\ (1 - r_{13.2}^2)\,]$$

$$R_{1.23}^2 = 1 - [\,1 - (0.4)^2\,]\,[\,1 - (0.76)^2\,] = 0.64$$

which is identical to the result obtained in Subsection 4.5.1 using eqs. 4.31 and 4.32.

Like the partial correlation, the *semipartial correlation coefficient* measures the correlation between $\mathbf{y}_j$ and $\mathbf{y}_k$ while controlling for the linear effect of other variables $\mathbf{y}_1, \mathbf{y}_2, \ldots$ and $\mathbf{y}_p$. The difference is in the denominator, which is the total variation in the response variable, i.e. the quantity [a+b+c+d] in Box 4.1. The formula for the first-order semipartial correlation coefficient is:

$$r_{1(2.3)} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1 - r_{23}^2}}$$

The value of $r_{1(2.3)}$ is 0 for the numerical example because [a] = 0. The *semipartial correlation* can also be calculated as the square root of the difference between two multiple determination coefficients:

$$r_{1(2.3)} = \sqrt{R_{1.23}^2 - R_{1.3}^2}$$

Because the latter equation does not specify the sign of the semipartial correlation coefficient, the previous equation must be used to obtain that sign, which is the same as the sign of the partial regression coefficient. In the Venn diagram of Box 4.1, $R_{1.23}^2$ is the union of the two ellipses or the quantity [a+b+c], whereas $R_{1.3}^2$ is the right-hand ellipse or the quantity [b+c], each of these quantities being divided by the total variation (total sum of squares) in the response variable, [a+b+c+d]. Hence $R_{1\,(2.3)}^2$ is ([a+b+c]−[b+c])/[a+b+c+d], or [a]/[a+b+c+d]. The semipartial correlation coefficient is especially useful in variation partitioning (Subsection 10.3.5) because it expresses all fractions of variation with respect to the same common denominator, which is the total sum of squares in the response variable [a+b+c+d].

Table 4.6 Main properties of the multiple (linear) correlation coefficient.

	Properties	Sections
1.	The multiple correlation coefficient measures the *intensity of the relationship* between a *response* variable and a *linear* combination of several *explanatory* variables.	4.5
2.	The square of the multiple correlation coefficient, called *coefficient of multiple determination*, measures the fraction of the variance of the response variable that is explained by a linear combination of the explanatory variables.	4.5
3.	The coefficient of multiple determination is the extension, to the multidimensional case, of the *coefficient of determination between two variables*.	4.5 and 10.3
4.	The multiple correlation coefficient can be computed from the matrix of correlations among *explanatory* variables and the vector of correlations between the *explanatory* and *response* variables.	4.5
5.	The multiple correlation coefficient can be computed from the determinant of the matrix of correlations among the *explanatory* variables and that of the matrix of correlations among all variables involved.	4.5
6.	The multiple correlation coefficient can be computed from the product of a series of *complements of coefficients of partial determination*.	4.5

Tables 4.6 and 4.7 summarize the main conclusions relative to the coefficients of multiple and partial correlation, respectively.

3 — Tests of statistical significance

In correlation analysis, the null hypothesis H_0 is usually that the correlation coefficient is equal to zero (i.e. independence of the descriptors). One can also test the hypothesis that ρ has some particular value other than zero. The general formula for testing correlation coefficients (for H_0: $\rho = 0$) is:

$$F = \frac{r_{jk}^2 / \nu_1}{(1 - r_{jk}^2) / \nu_2} \qquad \textbf{(4.39)}$$

with $\nu_1 = m$ and $\nu_2 = n - m - 1$, where m is the number of variables correlated to j. This F-statistic is compared to the critical value $F_{\alpha[\nu_1,\nu_2]}$. In the case of the simple correlation coefficient, where $m = 1$ (there is a single variable correlated to j), eq. 4.39 becomes eq. 4.12.

Table 4.7 Main properties of the partial (linear) correlation coefficient. One of these properties is discussed in a later chapter.

	Properties	Sections
1.	The partial and semipartial correlation coefficients measure the *intensity of the linear relationship* between two random variables while taking into account their relationships with other variables.	4.5
2.	The difference between the partial and semipartial correlation coefficients is in the denominator, which excludes the variation of the controlled variables in the partial correlation but not in the semipartial correlation.	4.5
3.	The partial correlation coefficient can be computed from the submatrix of correlations among the variables *in partial relationship* (first subset), the submatrix of variables that *influence* the first subset, and the submatrix of correlations between the *two subsets* of variables.	4.5
4.	The partial and semipartial correlation coefficients can be computed from the *coefficients of simple correlation* between all pairs of variables involved.	4.5
5.	The square of the partial correlation coefficient (*coefficient of partial determination*; name seldom used) measures the fraction of the total variance of each variable that is mutually explained by the other, the influence of some other variables being taken into account.	10.3

In regression analysis, the null hypothesis is that the coefficient of multiple determination (R^2) is zero. To test the *coefficient of multiple determination* R^2 and the *multiple correlation coefficient* R, the F-statistic is:

$$F = \frac{R^2_{1.2\ldots p} / \nu_1}{(1 - R^2_{1.2\ldots p}) / \nu_2} \qquad \textbf{(4.40)}$$

with $\nu_1 = m$ and $\nu_2 = n - m - 1$, where m is the number of explanatory variables; $m = p - 1$ in the notation of eq. 4.40.

Partial correlation coefficients are tested in the same way as coefficients of simple correlation (eq. 4.12 for the F-test and eq. 4.13 for the t-test, where $\nu = n - 2$), except that one additional degree of freedom is lost for each successive *order* of the coefficient, or each *covariable* in the model. For example, the number of degrees of freedom for $r_{jk.123}$ (third-order partial correlation coefficient) is $\nu = (n - 2) - 3 = n - 5$.

This is the same as counting $\nu = n - m - 1$, where m is the number of variables in the model besides j. For partial correlations, eqs. 4.12 and 4.13 become respectively:

$$F = \nu \frac{r^2_{jk.1\ldots p}}{1 - r^2_{jk.1\ldots p}} \quad \textbf{(4.12)} \qquad \text{and} \qquad t = \sqrt{\nu} \frac{r_{jk.1\ldots p}}{\sqrt{1 - r^2_{jk.1\ldots p}}} \quad \textbf{(4.13)}$$

The number of covariables will be called q in Subsections 10.3.5 and 11.1.7 which describe, respectively, the tests of significance in partial regression and partial canonical analysis. *Semipartial correlation coefficients* are tested using the same F-statistic as for partial correlations, as shown in Box 4.1. As usual (Sections 1.2 and 4.2), H_0 is tested either by comparing the computed statistic (F or t) to a critical value found in a table for a predetermined significance level α, or by computing the probability associated with the computed statistic.

4 — Causal modelling using correlations

In the ecological literature, correlation coefficients are often interpreted in terms of causal relationships among descriptors. That should never be done when the only information available is that provided by the correlation coefficients themselves.

Causality

In statistics, "causality" refers to the hypothesis that changes occurring in one variable cause changes in another variable; *causality resides in the hypotheses only.* Within the framework of a given sampling design (i.e. spatial, temporal, or experimental) where variation is controlled, data are said to support the causality hypothesis if a significant portion of the variation in **b** is explained by changes taking place in **a**. If the relationship is assumed to be linear, a significant linear correlation coefficient is interpreted as supporting the hypothesis of linear causation.

Let us consider the simple case of three linearly related variables $\mathbf{y}_1$, $\mathbf{y}_2$, and $\mathbf{y}_3$. In the following paragraphs, these variables will be noted **a**, **b**, and **c** for simplicity. A simple form of causal modelling is obtained by looking at the simple and partial correlation coefficients between these variables, following the pioneering work of De Neufville & Stafford (1971). One basic condition must be fulfilled for such a model to encompass the three variables; it is that at least two of the simple correlations be significantly different from zero. Under the assumption of linear relationships among variables, these two coefficients provide statistical support for two "causal arrows".

Causal model

There are four elementary models describing the possible interactions among three variables (Fig. 4.11), each with possible permutations of **a**, **b** and **c**, for a total of 18 distinguishable models. These four elementary *causal models* show how difficult it is to interpret correlation matrices, especially when several ecological descriptors are interacting in complex ways. Partial correlations may be used to elucidate the relationships among descriptors. However, the choice of a causal model always requires hypotheses, or else the input of external ecological information. When it is possible, from a priori information or ecological hypotheses, to specify the causal

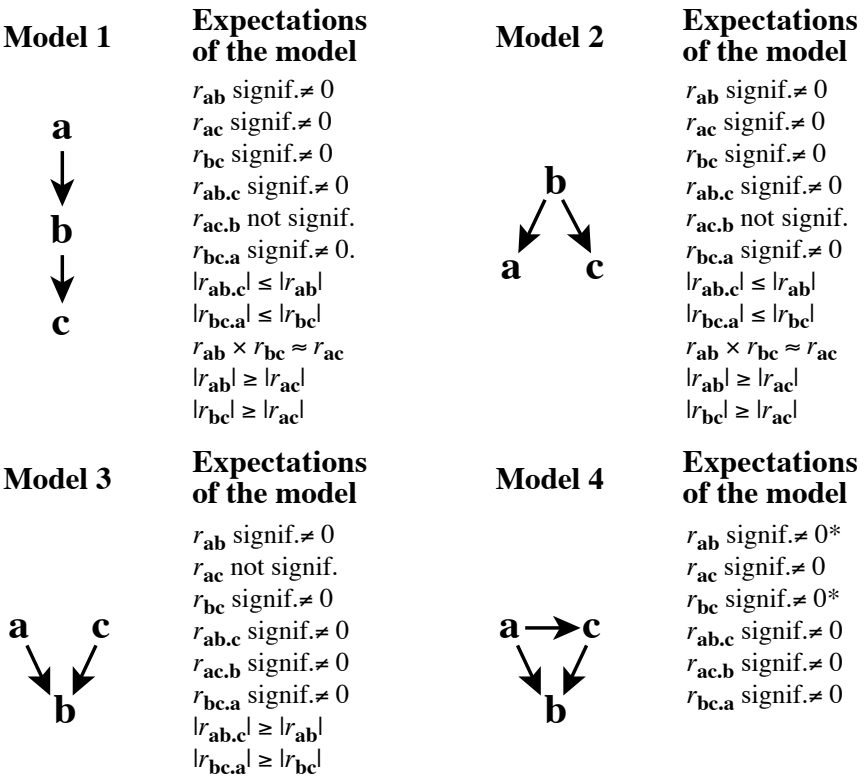


Figure 4.11 Predictions of the four possible models of causal relationships involving three variables, in terms of the expected values for the simple and partial linear correlation coefficients. 'r_{ab} signif.≠ 0' means that, under the model, the correlation must be significantly different from zero. 'r_{ab} not signif.' means that the correlation is not necessarily significantly different from zero at the pre-selected significance level. * Model 4 holds even if one, *but only one*, of these two simple correlation coefficients is not significant. Adapted from Legendre (1993).

ordering among descriptors, path analysis (Section 10.4) may be used to assess the correspondence between the data (i.e. correlations) and causal models. Note again that a causal model may never be derived from a correlation matrix, whereas a causal model is required to interpret a correlation matrix in terms of causality.

In Fig. 4.11, model 1 describes a *causal chain*, with six possible permutations of **a**, **b** and **c**, and model 2 is the *double effect* model with three distinguishable permutations: each of the three variables may be at the origin of the two arrows. Model 3 is the *double cause* model, with three distinguishable permutations. Model 4 describes a *triangular relationship* with six possible permutations; it may be seen as a combination of models 1 and 2 or 1 and 3. The direct and indirect effects implied in

model 4 may be further analysed using path analysis (Section 10.4), which requires precise hypotheses about arrow directions.

In Fig. 4.11, the predictions of the four models were obtained by numerical simulations. Examining model 1 in some detail illustrates how the "expectations of the model" can also be derived analytically.

• Significance of the simple correlations. Obviously, $r_{\mathbf{ab}}$ and $r_{\mathbf{bc}}$ must be significantly different from zero for the model to hold. The model can accommodate $r_{\mathbf{ac}}$ being significant or not, although the value of $r_{\mathbf{ac}}$ should always be different from zero since $r_{\mathbf{ac}} = r_{\mathbf{ab}} r_{\mathbf{bc}}$.

• Significance of the partial correlations. The condition $r_{\mathbf{ac}} = r_{\mathbf{ab}} r_{\mathbf{bc}}$ implies that $r_{\mathbf{ac}} - r_{\mathbf{ab}} r_{\mathbf{bc}} = 0$ or, in other words (eq. 4.36), $r_{\mathbf{ac.b}} = 0$. For the model to hold, partial correlations $r_{\mathbf{ab.c}}$ and $r_{\mathbf{bc.a}}$ must be significantly different from 0. Indeed, $r_{\mathbf{ab.c}}$ being equal to zero would mean that $r_{\mathbf{ab}} = r_{\mathbf{ac}} r_{\mathbf{bc}}$, which would imply that **c** is in the centre of the sequence; this is not the case in the model as specified, where **b** is in the centre. The same reasoning explains the relationship $r_{\mathbf{bc.a}} \neq 0$.

• Comparison of simple correlation values. Since correlation coefficients are smaller than or equal to 1 in absolute value, the relationship $r_{\mathbf{ac}} = r_{\mathbf{ab}} r_{\mathbf{bc}}$ implies that $|r_{\mathbf{ab}}| \geq |r_{\mathbf{ac}}|$ and $|r_{\mathbf{bc}}| \geq |r_{\mathbf{ac}}|$.

• Comparison of partial correlation values. Consider the partial correlation formula for $r_{\mathbf{ab.c}}$ (eq. 4.36). Is it true that $|r_{\mathbf{ab.c}}| \leq |r_{\mathbf{ab}}|$? The relationship $r_{\mathbf{ac}} = r_{\mathbf{ab}} r_{\mathbf{bc}}$ allows one to replace $r_{\mathbf{ac}}$ by $r_{\mathbf{ab}} r_{\mathbf{bc}}$ in that equation. After a few lines of algebra, the following inequality

$$|r_{\mathbf{ab.c}}| = \frac{|r_{\mathbf{ab}}| \, [1 - r_{\mathbf{bc}}^2]}{\sqrt{[1 - r_{\mathbf{ab}}^2 r_{\mathbf{bc}}^2] \, [1 - r_{\mathbf{bc}}^2]}} \leq |r_{\mathbf{ab}}|$$

leads to the relationship $r^2_{\mathbf{bc}} (1 - r^2_{\mathbf{ab}}) \geq 0$, which is true in all cases because $r_{\mathbf{bc}} \neq 0$ and $|r_{\mathbf{ab}}| \leq 1$. This also shows that $r_{\mathbf{ab.c}} = r_{\mathbf{ab}}$ only when $r_{\mathbf{ab}} = 1$. The same method can be used to demonstrate that $|r_{\mathbf{bc.a}}| \leq |r_{\mathbf{bc}}|$.

The model predictions in Fig. 4.11 show that it is not possible to distinguish between models 1 and 2 from the correlation coefficients alone: these two models differ only in their hypotheses (arrow directions). Their key common characteristic is the non-significance of the partial correlation $r_{\mathbf{ac.b}}$. Model 3 is distinct in the fact that $r_{\mathbf{ac}}$ is not significant and that the partial correlations are, in absolute values, larger than or equal to the corresponding simple correlations, whereas they are smaller in models 1 and 2. For model 4, some of the predictions depend on the signs of the effects depicted by the arrows; for example, the three partial correlations may be larger or smaller, in absolute values, than the simple correlations. Model 4 may apply even if one, but only one, of the two simple correlations, $r_{\mathbf{ab}}$ or $r_{\mathbf{bc}}$, is not significant. When n

is small, the tests may not have enough power to evidence the significance of the relationships and, as a consequence, evidence may be lacking to support a model.

Numerical example. The simple example already used for multiple and partial correlations illustrates here the problem inherent to all correlation matrices, i.e. that it is never possible to interpret correlations *per se* in terms of causal relationships. In the following matrix, the upper triangle contains the coefficients of simple correlation whereas the lower triangle contains the partial correlation coefficients:

$$\begin{bmatrix} 1 & 0.4 & 0.8 \\ 0 & 1 & 0.5 \\ 0.76 & 0.33 & 1 \end{bmatrix}$$

It may have seemed that descriptors $\mathbf{y}_1$ and $\mathbf{y}_2$ were correlated ($r_{12} = 0.4$), but the first-order partial correlation coefficient $r_{12.3} = 0$ shows that this is not the case. The predictions of models 1 and 2 in Fig. 4.11, with $\mathbf{a} = \mathbf{y}_1$, $\mathbf{b} = \mathbf{y}_3$ and $\mathbf{c} = \mathbf{y}_2$, are in agreement with these results. In the absence of external information or ecological hypotheses, there is no way of determining which pattern of causal relationships, model 1 or model 2, actually fits this correlation matrix.

Ecological application 4.5

Bach *et al*. (1992) analysed a 28-month long time series (weekly sampling, $n = 122$) of eel catches (*Anguilla anguilla*) in the Thau marine lagoon in southern France. Fixed gears called 'capêchades', composed of three funnel nets (6-mm mesh) and an enclosure, were used near the shore in less than 1.5 m of water. In the deeper parts of the lagoon, other types of gears were used: heavier assemblages of funnel nets with larger mesh sizes, called 'brandines', 'triangles' and 'gangui', as well as longlines. Various hypotheses were stated by the authors and tested using partial correlation analysis and path analysis. These concerned the influence of environmental variables, including air temperature as a proxy for seasons, on the behaviour of fish and fishermen, and their effects on landings. Coefficients of linear correlation reported in the paper are used here to study the relationships among air temperature, fishing effort, and landings, for the catches by 'capêchade' (Fig 4.12). The analysis in the paper was more complex; it also considered the effects of wind and lunar phases. Linearity of the relationships was checked. The correlation coefficients are consistent with a type-4 model stating that both effort and temperature affect the landings (temperature increases eel metabolism and thus their activity and catchability) and that the effort, represented by the number of active 'capêchade' fishermen, is affected by seasonality (lower effort at high temperature, 'capêchades' being not much used from August to October). Interesting is the non-significant simple linear correlation between temperature and catches. The partial correlations indicate that this simple correlation corresponds to two effects of temperature on catches that are both significant but of opposite signs: a positive partial correlation of temperature on catches and a negative one of temperature on effort. In the paper of Bach *et al*., partial correlation analysis was used as a first screen to eliminate variables that clearly did not influence catches. Path analysis (Section 10.4) was then used to study the direct and indirect effects of the potentially explanatory variables on catches.

Partial correlations do not provide the same information as path analysis (Section 10.4). On the one hand, partial correlations, like partial regression coefficients (Subsection 10.3.3), indicate whether a given variable has some unique (linear)

	Eel catches	Fishing effort	Air temperature
Eel catches	----	$r = 0.730$***	$r = 0.096$ (N.S.)
Fishing effort	partial r = 0.763***	----	$r = -0.180$*
Air temperature	partial r = 0.338***	partial r = –0.368***	----

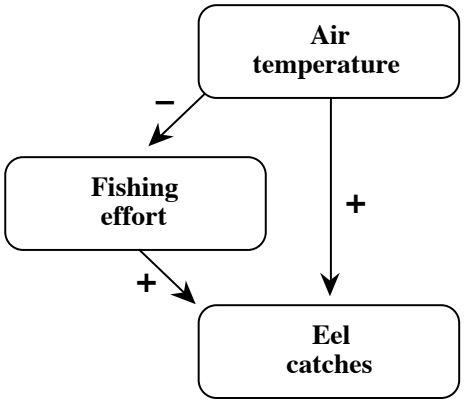


Figure 4.12 Left: simple and partial correlations among temperature, fishing effort, and eel catches using the 'capêchade' fishing gear, from Bach *et al*. (1992). Right: causal model supported by the data. The signs of the partial correlation coefficients are shown along the arrows. *: $0.05 \geq p > 0.01$; ***: $p \leq 0.001$; N.S.: non-significant correlation at significance level $\alpha = 0.05$.

relationship with some other variable, after the linear effects of all the other variables in the model have been taken into account. In path analysis on the other hand, one is mostly interested in partitioning the relationship between predictor (explanatory, independent) and criterion (response, dependent) variables into direct and indirect components.

The above discussion was based on linear correlation coefficients. Advantages of the linear model include ease of computation and simplicity of interpretation. However, environmental processes are not necessarily linear. This is why linearity must be checked, not only assumed, before embarking in this type of computation. When the variables are not linearly related, two choices are open: either proceed with non-linear statistics (nonparametric simple and partial correlation coefficients, in particular, are available and may be used in this type of calculation), or linearize the relationships that seem promising. Monotonic relationships, identified in scatter diagrams, can often be linearized by applying the transformations of Section 1.5 to one or both variables. There is no 'cheating' involved in doing that; either a monotonic relationship exists, and linearizing transformations allow one to apply linear statistics to the data; or there is no monotonic relationship, and no amount of transformation will ever create one.

Simple causal modelling, as presented in this subsection, may be used in two different types of circumstances. A first, common application is exploratory analysis, which is performed when 'weak' ecological hypotheses only can be formulated. What this means is the following: in many studies, a large number of causal hypotheses may

be formulated *a priori*, some being contradictory, because the processes at work in ecosystems are too numerous for ecologists to decide which ones are dominant under given circumstances. So, insofar as each of the models derived from ecological theory can be translated into hypothesized correlation coefficients, partial correlation analysis may be used to clear away those hypotheses that are not consistent with the data and to keep only those that look promising for further analysis. Considering three variables, for instance, one may look at the set of simple and partial correlation coefficients and decide which of the four models of Fig. 4.11 are not consistent with the data. Alternatively, when the ecosystem is better understood, one may wish to test a single set of hypotheses (i.e. a single model), to the exclusion of all others. With three variables, this would mean testing only one of the models of Fig. 4.11, to the exclusion of all others, and deciding whether the data are consistent, or not, with that model.

Several correlation coefficients are tested in each panel of Fig. 4.11. Three simultaneous tests are performed for the simple correlation coefficients and three for the partial correlation coefficients. In order to determine whether such results could have been obtained by chance alone, some kind of global test of significance, or correction, must be performed (eq. 4.14; Box 1.3).

The simple form of modelling described here may be extended beyond the frame of linear modelling, as long as formulas exist for computing partial relationships. Examples are the partial nonparametric correlation coefficient (partial Kendall τ, eq. 5.9) and the partial Mantel statistic (Subsection 10.5.2).

4.6 Tests of normality and multinormality

Testing the normality of empirical distributions is an important concern for ecologists who want to use linear models for analysing their data. Tests of normality are carried out in two types of circumstances. On the one hand, many tests of statistical significance, including those described in the present chapter, require the empirical data to be drawn from normally distributed populations. On the other hand, the linear methods of multivariate data analysis discussed in Chapters 9, 10, and 11 do summarize data in more informative ways if their underlying distributions are multinormal — or at least are not markedly skewed, as discussed below. Estimating the skewness and testing the normality of empirical variables is thus an important initial step in the analysis of a data set. Variables that are not normally distributed may be subjected to normalizing transformations (Section 1.5). The historical development of the tests of normality has been reviewed by D'Agostino (1982) and by Dutilleul & Legendre (1992).

The problem may first be approached by plotting frequency histograms of empirical variables. Looking at these plots immediately identifies distributions that have several modes, for instance, or are obviously too skewed or too 'flat' or 'peaked' to have possibly been drawn from normally distributed populations.

Next, for unimodal distributions, one may examine the skewness and kurtosis parameters. The first centred moment of a distribution is $m_1 = 0$ and the second is the variance, $m_2 = s_x^2$ (unbiased estimator: eq. 4.3). The unbiased estimator of the third centred moment is:

$$m_3 = \frac{n \sum (x_i - \bar{x})^3}{(n-1)\,(n-2)}$$

Skewness

Skewness (α_3) is a measure of asymmetry. It is estimated as the third moment of the distribution divided by the cube of the standard deviation:

$$\alpha_3 = m_3 / s_x^3 \qquad \textbf{(4.41)}$$

Skewness is 0 for a symmetric distribution like the normal distribution. Positive skewness corresponds to a frequency distribution with a longer tail to the right than to the left, whereas a distribution with a longer tail to the left shows negative skewness. The unbiased estimator of the fourth moment of a distribution is:

$$m_4 = \frac{n\,(n+1) \sum (x_i - \bar{x})^4 - 3\,(n-1)\,(\sum (x_i - \bar{x})^2)^2}{(n-1)\,(n-2)\,(n-3)}$$

Kurtosis

Kurtosis (α_4) is a measure of flatness or peakedness of a distribution. It is estimated as the fourth moment divided by the standard deviation to the power 4:

$$\alpha_4 = m_4 / s_x^4 \qquad \textbf{(4.42)}$$

The kurtosis of a normal distribution is $\alpha_4 = 0$. Distributions flatter than the normal distribution ('platycurtic') have negative values for α_4 whereas distributions that have more observations around the mean than the normal distribution have positive values for α_4, indicating that they are 'leptokurtic' which means more 'peaked'. The value of α_4 for a uniform (flat, rectangular) distribution is -1.2.

Although tests of significance have been developed for skewness and kurtosis, they are not used any longer because more powerful tests of normality are now available. For the same reason, testing the goodness-of-fit of an empirical frequency distribution to a normal distribution with same mean and variance (as in Fig 4.13a) using a chi-square test is no longer in fashion because it is not very sensitive to departures from normality (Stephens, 1974; D'Agostino, 1982), even though it may still be presented in some texts of biological statistics as an acceptable procedure. The main problem is that it does not take into account the ordering of classes of the two frequency distributions that are being compared. This explains why the main statistical packages do not use it, but propose instead one or the other (or both) procedure described below.

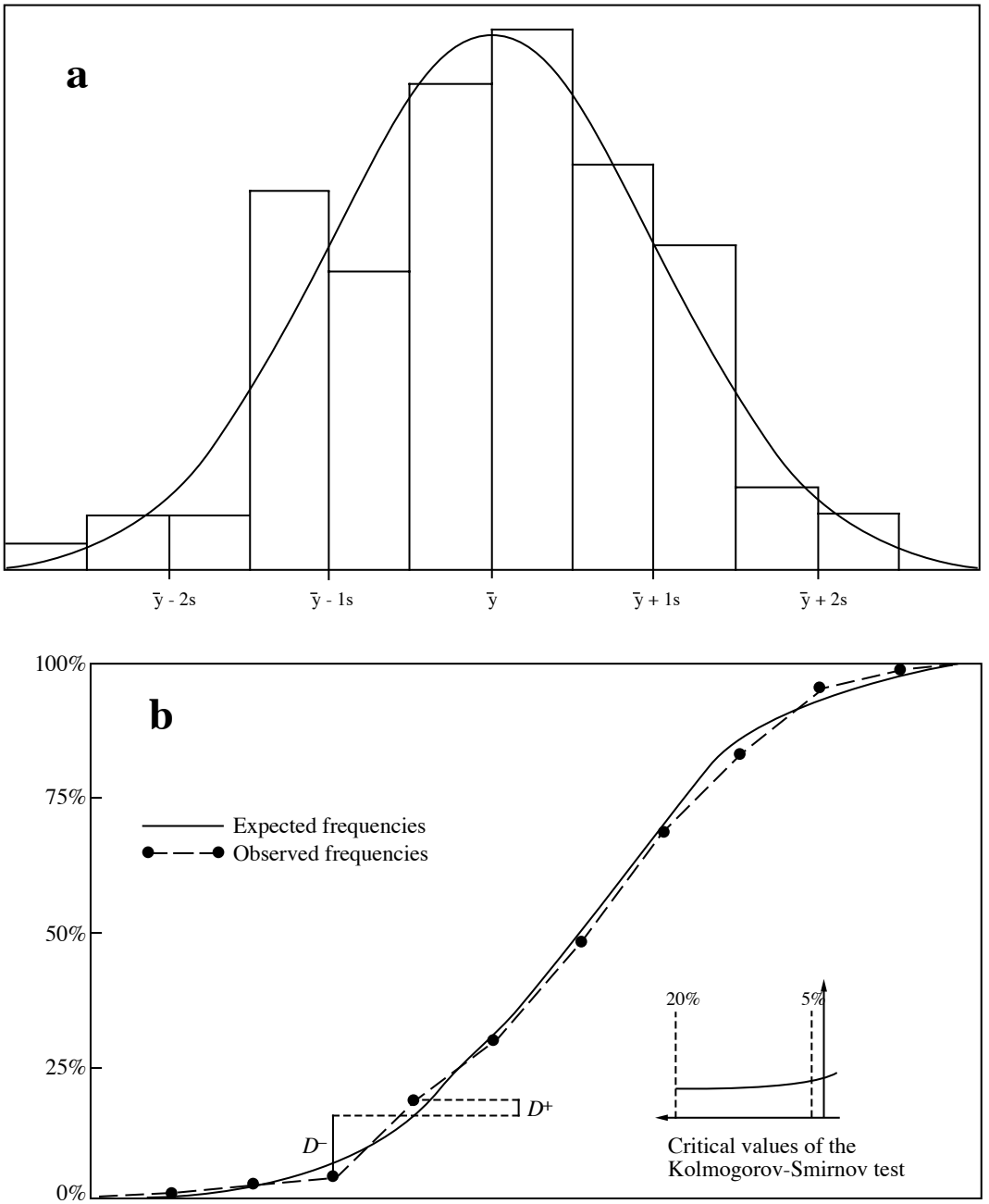


Figure 4.13 Numerical example with $n = 100$. (a) Frequency distribution and fitted theoretical normal curve, (b) relative cumulative frequencies and Kolmogorov-Smirnov test of goodness-of-fit, showing that the maximum deviation $D = 0.032$ is too small in this case to reject the hypothesis of normality.

One of the widely used tests of normality is the Kolmogorov-Smirnov test of goodness-of-fit. In Fig. 4.13b, the same data as in Fig. 4.13a are plotted as a cumulative frequency distribution. The cumulative theoretical normal distribution is also plotted on the same graph; it can easily be obtained from a published table, or by requesting in a statistical package the normal probability values corresponding to the relative cumulative frequencies (function ***pnorm()*** in R). One looks for the largest deviation D between the cumulative empirical relative frequency distribution and the cumulative theoretical normal distribution. If D is larger than or equal to the critical value in the table, for a given number of observations n and significance level α, the hypothesis of normality is rejected.

K-S test

The Kolmogorov-Smirnov (K-S) test of goodness-of-fit is especially interesting for small sample sizes because it does not require to lump the data into classes. When they are divided into classes, the empirical data are discontinuous and their cumulative distribution is a step-function, whereas the theoretical normal distribution to which they are compared is a continuous function. D is then formally defined as the maximum of D^- and D^+, where D^- is the maximum difference computed just before a data value and D^+ is the maximum difference computed at the data value (i.e. at the top of each step of the cumulative empirical step-function). A detailed numerical example of the procedure is presented by Sokal & Rohlf (1995).

Standard Kolmogorov-Smirnov tables for the comparison of two samples, where the distribution functions are completely specified (i.e. the mean and standard deviation are stated by hypothesis), are not appropriate for testing the normality of *empirical data* since the mean and standard deviation of the reference normal distribution must then be estimated from the observed data; critical values given in these tables are systematically too large, and thus lead too often to not rejecting the null hypothesis of normality. Corrected critical values for testing whether a set of observations is drawn from a normal population, that are valid for stated probabilities of type I error, have been computed by Lilliefors (1967) and, with additional corrections based on larger Monte Carlo simulations, by Stephens (1974). The same paper by Stephens evaluates other statistics to perform tests of normality, such as Cramér-von Mises W^2 and Anderson-Darling A^2 which, like D, are based on the empirical cumulative distribution function (only the statistics differ); it proposes corrections where needed for the situation where the mean and variance of the reference normal distribution are unknown and are thus estimated from the data.

Normal probability plot

The second widely used test of normality is due to Shapiro & Wilk (1965). It is based on an older graphical technique that will be described first. This technique, called *normal probability plotting*, was developed as an informal way of assessing deviations from normality. The objective is to plot the data in such a way that, if they come from a normally distributed population, they will fall along a straight line. Deviations from a straight line may be used as indication of the type of non-normality. In these plots, the values along the abscissa are either the observed or the standardized data (in which case the values are transformed to standard deviation units), while the ordinate is the percent cumulative frequency value of each point plotted on a normal

probability scale. Sokal & Rohlf (1995) give computation details. Figure 4.14 shows the same data as in Fig 4.13a, which are divided into classes, plotted on normal probability paper. The same type of plot could also be produced for the raw data, not grouped into classes. For each point, the *upper limit* of a class is used as the abscissa, while the ordinate is the percent cumulative frequency (or the cumulative percentage) of that class. Perfectly normal data would fall on a straight line passing through the point ($\bar{y}$, 50%). A straight line is fitted trough the points, using reference points based on the mean and variance of the empirical data (see the legend of Fig. 4.14); deviations from that line indicate non-normality. Alternatively, a straight line may be fitted through the points, either by eye or by regression; the mean of the distribution may be estimated as the abscissa value that has an ordinate value of 50% on that line. D'Agostino (1982) gives examples illustrating how deviations from linearity in such plots indicate the degree and type of non-normality of the data.

Shapiro-Wilk test

Shapiro & Wilk (1965) proposed to quantify the information in normal probability plots using a statistic called 'analysis of variance *W*', which they defined as the *F*-ratio of the estimated variance obtained from the weighted least-squares of the slope of the straight line (numerator) to the variance of the sample data (denominator). The statistic is used to assess the goodness of the linear fit:

$$W = \left(\sum_{i=1}^{n} w_i x_i \right)^2 \Big/ \sum_{i=1}^{n} (x_i - \bar{x})^2 \qquad \textbf{(4.43)}$$

where the x_i are the ordered observations ($x_1 \le x_2 \le \ldots \le x_n$) and coefficients w_i are optimal weights for a population assumed to be normally distributed. Statistic *W* may be viewed as the square of the correlation coefficient (i.e. the coefficient of determination) between the abscissa and ordinate of the normal probability plot described above. Large values of *W* indicate normality (points lying along a straight line give r^2 close to 1), whereas small values indicate lack of normality. Shapiro & Wilk did provide critical values of *W* for sample sizes up to 50. D'Agostino (1971, 1972) and Royston (1982a, b, c) proposed modifications to the *W* formula (better estimates of the weights w_i), which extend its application to much larger sample sizes. Extensive simulation studies have shown that *W* is a sensitive *omnibus* test statistic, meaning that it has good power properties over a wide range of non-normal distribution types and sample sizes.

The Shapiro-Wilk test is available in the ***shapiro.test()*** function of the R STATS package. Five other functions are available in the NORTEST package to carry out tests of normality, including function ***lillie.test()*** for the Lilliefors (1967) K-S test using Stephens' (1974) corrections. Which of these tests is best? Reviewing the studies on the power of tests of normality published during the previous 25 years, D'Agostino (1982) concluded that the best *omnibus* tests are the Shapiro-Wilk *W*-test and a modification by Stephens (1974) of the Anderson-Darling A^2-test mentioned above (***ad.test()*** function in NORTEST). In a Monte Carlo study involving autocorrelated data (Section 1.1), however, Dutilleul & Legendre (1992) showed (1) that, for moderate

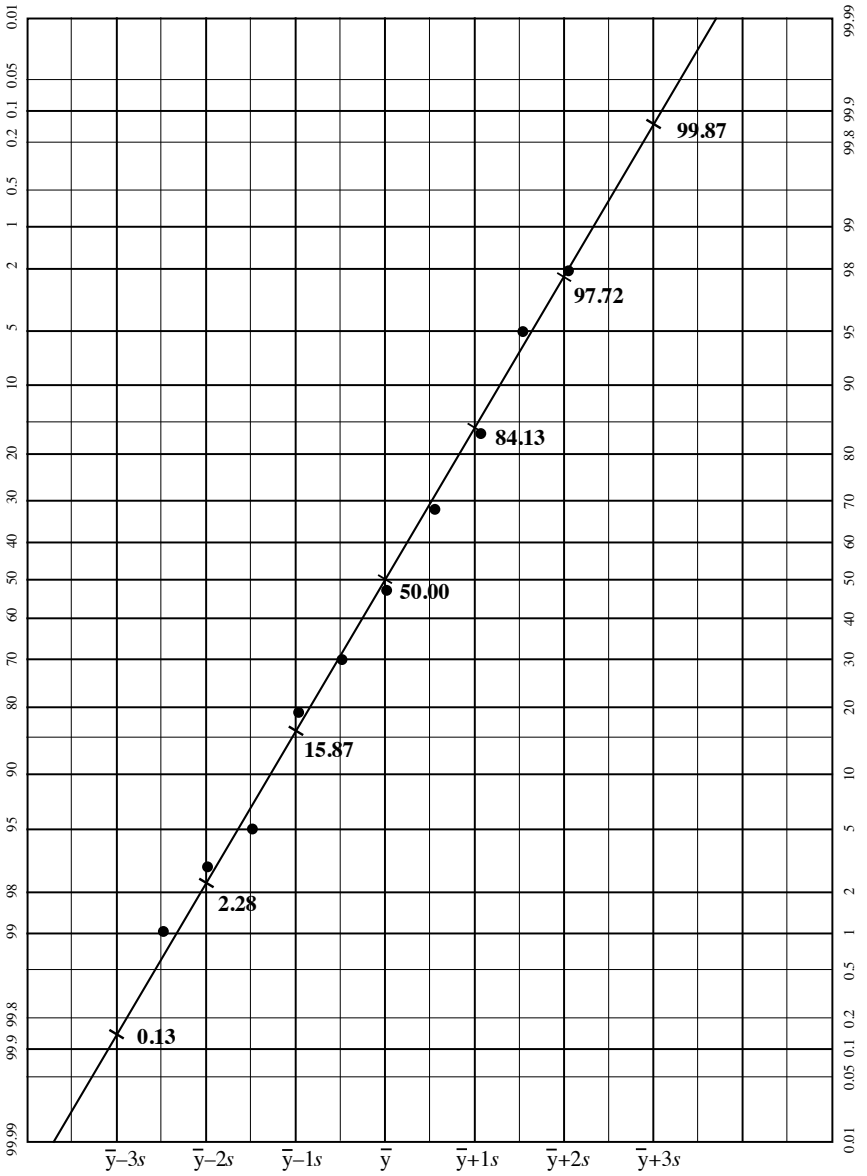


Figure 4.14 The cumulative percentages of data in Fig. 4.13a are plotted here on normal probability paper (probit transformation) as a function of the upper limits of classes. Cumulative percentiles are indicated on the right-hand side of the graph. The last data value cannot be plotted on this graph because its cumulated percentage value is 100. The diagonal line represents the theoretical cumulative normal distribution with same mean and variance as the data. This line is positioned on the graph using reference values of the cumulative normal distribution, for example 0.13% at $\bar{y} - 3s$ and 99.87% at $\bar{y} + 3s$, and it passes through the point ($\bar{y}$, 50%). This graph contains exactly the same information as Fig. 4.13b; the difference lies in the scale of the ordinate.

sample sizes, both the *D*-test and the *W*-test were too liberal (in an asymmetric way) for high positive ($\rho > 0.4$) and very high negative ($\rho < -0.8$) values of autocorrelation along time series and for high positive values of spatial autocorrelation ($\rho > 0.2$), and (2) that, overall, the Kolmogorov-Smirnov *D*-test was more robust against autocorrelation than the Shapiro-Wilk *W*-test, whatever the sign of the first-order autocorrelation.

As stated at the beginning of this section, ecologists must absolutely check the normality of data only when they are planning to use parametric statistical tests that assume normality of the distributions; permutation tests (Section 1.2) can be used with non-normal data. Most methods presented in this book, including clustering and ordination techniques, do not require statistical testing and hence may be applied to non-normal data. With many of these methods, however, ecological structures emerge more clearly when the data do not present strong asymmetry; this is the case, for example, with principal component analysis. Since normal data are not skewed (coefficient $\alpha_3 = 0$), testing the normality of data is also testing for asymmetry; normalizing transformations, applied to data with unimodal distributions, reduce or eliminate asymmetry. So, with multidimensional data, it is recommended to check at least the skewness of the variables one by one.

Some tests of significance require that the data be multinormal (Section 4.3). Normality of the *p* individual variables can easily be tested as described above. In a multivariate situation, however, showing that each variable does not significantly depart from normality does not demonstrate that the multivariate data set is multinormal although, in many instances, this is the best that researchers can do.

Test of multinormality

Dagnelie (1975) proposed an elegant way of testing the multivariate normality of a set of multivariate observations. The method is based on the *Mahalanobis generalized distance* (D_5; Section 7.4, eq. 7.38) described in Chapter 7. Generalized distances are computed, in the multidimensional space, between each object and the multidimensional mean of all objects. The distance between object $\mathbf{x}_i$ and the mean point $\bar{\mathbf{x}}$ is computed as:

$$D(\mathbf{x}_i, \bar{\mathbf{x}}) = \sqrt{[y - \bar{y}]_i \mathbf{S}^{-1} [y - \bar{y}]_i'} \qquad \textbf{(4.44)}$$

where $[y - \bar{y}]_i$ is the vector corresponding to object $\mathbf{x}_i$ in the matrix of centred data and **S** is the multivariate dispersion matrix (Section 4.1). For standardized variables $z_{ij} = (y_{ij} - \bar{y}_j) / s_j$, eq. 4.44 becomes:

$$D(\mathbf{x}_i, \bar{\mathbf{x}}) = \sqrt{\mathbf{z}_i \mathbf{R}^{-1} \mathbf{z}_i'} \qquad \textbf{(4.45)}$$

where **R** is the correlation matrix. Dagnelie's approach is that, for multinormal data, the generalized distances should be normally distributed. He suggested to do a visual examination of the cumulative frequency distribution as in Fig. 4.14. Actually, the generalized distances can be subjected to a Shapiro-Wilk test of normality, whose conclusions are applied to the multinormality of the original multivariate data.

The Dagnelie test of multivariate normality based on the Shapiro-Wilk test of normality of Mahalanobis generalized distances is invalid for univariate data (type I error rate too high). Numerical simulations by D. Borcard (personal communication) showed that the test had correct levels of type I error for values of n between $3p$ and $7.5p$, where p is the number of variables in the data table (simulations with $1 \leq p \leq 50$). Outside that range of n values, the results were too liberal, meaning that the test rejected too often the null hypothesis of normality. For $p = 2$, the simulations showed the test to be valid for $6 \leq n \leq 11$. If H_0 is not rejected in a situation where the test is too liberal, the result is trustworthy.

4.7 Software

Functions for all operations described in this chapter are available in the R language.

1. Covariance matrices are computed using functions ***var()*** and ***cov()*** of the STATS package; correlation matrices are computed by ***cor()***. The F-test comparing two variances is carried out by ***var.test()*** and correlation coefficients are tested using ***cor.test()*** in STATS.

2. Eigenanalysis is computed by ***eigen()*** in STATS.

3. Partial correlations are computed by function ***partial.cor()*** of the RCMDR package.

4. Tests of normality are computed using ***shapiro.test()*** in STATS, ***lillie.test()*** in NORTEST, and ***ad.test()*** in NORTEST. Function ***qqnorm()*** of STATS produces normal quantile-quantile plots like Fig. 4.14.

5. Function ***pnorm()*** in STATS computes p-values for the normal distribution, ***pf()*** for the F-distribution, ***pt()*** for the Student t-distribution, ***pchisq()*** for the chi-square distribution, and so on for other statistical distributions.

Commercial statistical packages, as well as S-PLUS® and MATLAB®, also provide functions for these calculations.