2 Descriptive Statistics

2.1 Graphical Representations

a **Scatter Plots.** What we detemine with our eyes is almost always put down in two dimensions, typically on paper or on a screen. Therefore, graphical representations are also almost always two dimensional. The joint distribution of two variables can be represented in a scatter plot in a natural way. Fig. 2.1.a shows the log lengths and widths of the sepal leaves of the 50 Iris setosa plants in the Iris flower example.

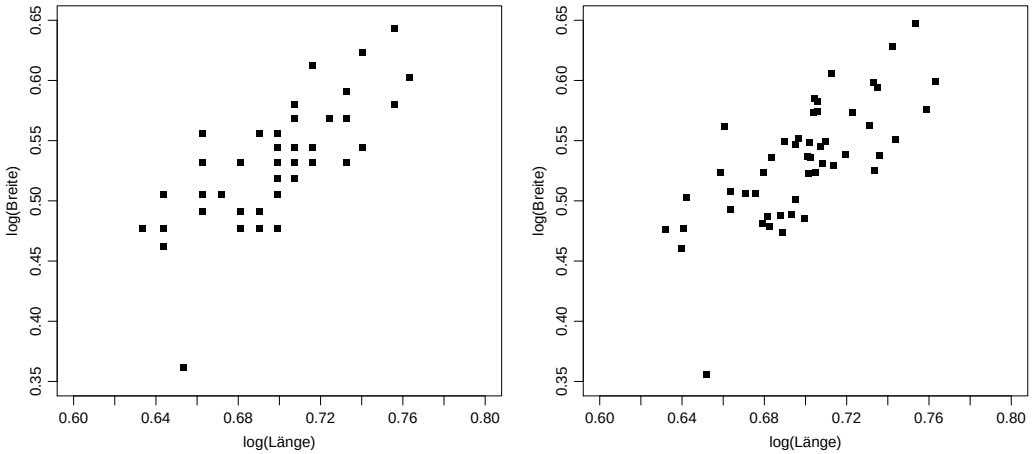

Figure 2.1.a: Log Lengths and Widths of the Sepal Leaves of Iris setosa, Without and With "Jittering"

If we use the usual function for a scatter plot (left picture in the figure) it is not apparent that some points lie on top of each other because of rounding of observations. One way to make such **multiple points** visible is to "jitter" the observations, which means that we add a small, random number. This is done for the right panel.

b With some effort we can exploit **three dimensions** by continuously altering the viewpoint for the two dimensional projection of a three-dimensional pattern, or by other forms of producing a three dimensional impression. In this way, three variables can be represented jointly. For four variables or more we have to come up with something else.

c There are many ideas about representing several variables in two dimensions (without animation, i.e. statically). Several are described in Stahel (2002), Ch. 3.6. There are whole books full of elaborate methods for such representations.

- Three books that link the useful with the artistic come from E. Tufte (1983, 1990, 1997).
- W. Cleveland studied the effectivity of statistical graphical representations and developed from this a new style of graphical representation, which has the name "trellis". In his books Cleveland (1993) and Cleveland (1994) he presents many useful applications of this representation for explorative data analysis and the presentation of statistical results.

d **Scatter Plot Matrix.** A basic, simple type of representing more than two variables is to summarize the scatter plots of all possible pairs of variables into a matrix (Fig. 2.1.d).

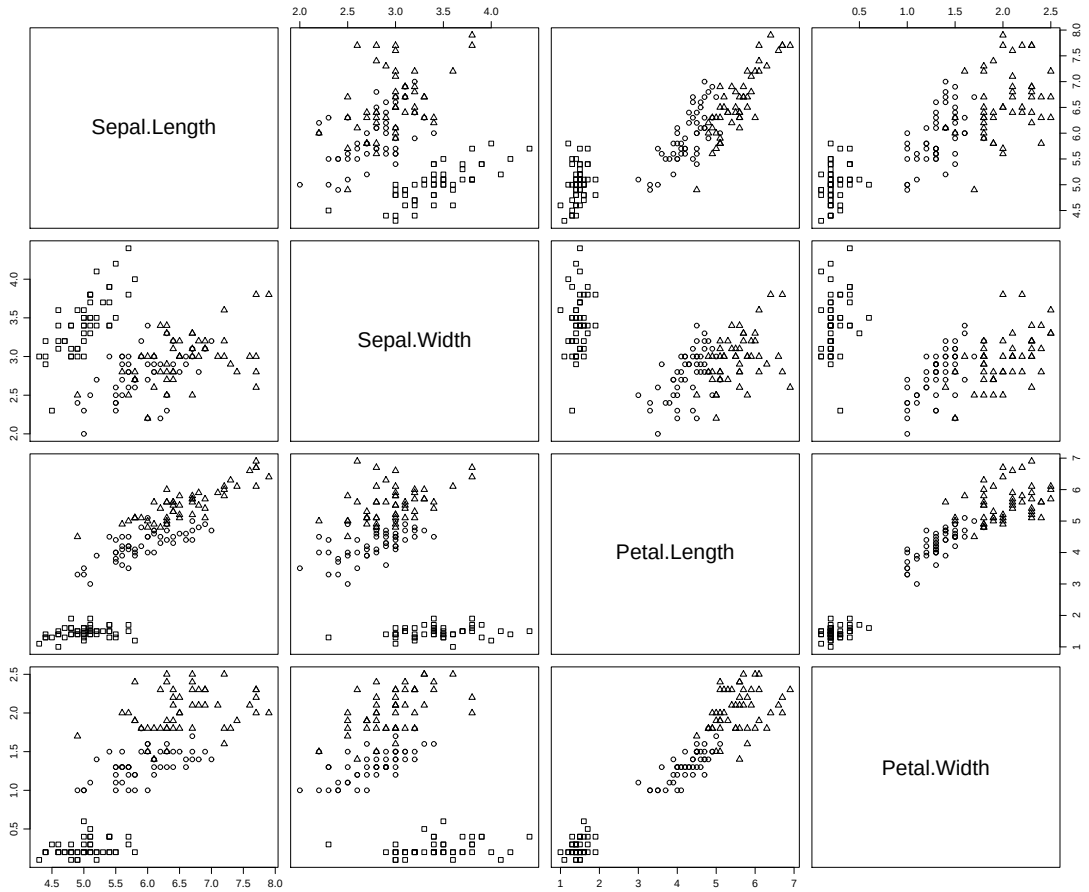

Figure 2.1.d: Scatter Plot Matrix of the Iris Data. The various symbols stand for the three different iris species.

In each row of this matrix, one variable is shown in the vertical direction, while all others are used in the row as horizontal axes. In each column of the matrix, one variable appears as the horizontal axis and all others as vertical axes. Since in the "diagonals" the variables are shown against themselves, these fields are to be used another way. In Fig. 2.1.d the names of the variables are written here. Often, histograms of the variables are shown there.

Because nothing really new comes out of switching the axes, we can limit ourselves to one of the two possible scatter plots per variable pair. For this reason, only the bottom half of the matrix is frequently shown.

e **Coplot.** The scatter plots represent only two dimensional distributions of the observations. A so-called **Coplot** can show complicated relationships between four variables. It is based on first classifying two variables into, for example, 6 overlapping regions with equal numbers of observations, and the graph is split into 6×6 partial areas. In the partial area $[h, \ell]$ only the data that belongs to the corresponding combination of classes of the two variables is used. For this data a scatterplot of the two remaining variables is drawn.

In Fig. 2.1.e the length and width of the sepal leaves of the iris flowers form the axes of the scatterplot, while for the allocations the length of the petal leaves and the plant species are used. The latter variable is already a categorical variable and does not need to be divided into classes.

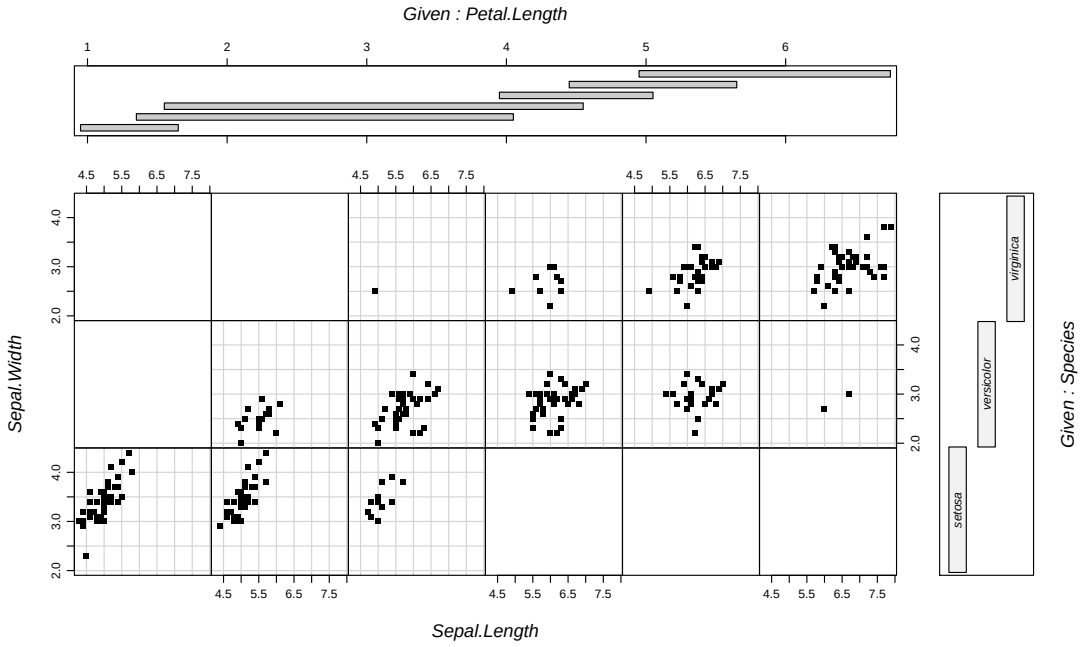

Figure 2.1.e: Coplot of the Iris Data

2.2 Symbols

a **Scatterplots with Symbols.** In a scatterplot of two variables, more variables can be represented with various types of symbols.

▷ In Fig. 2.2.a two variables that characterize the soil form the coordinates of the scatterplots, while the symbols reflect the number of the found specimens of five important plant species. ◁

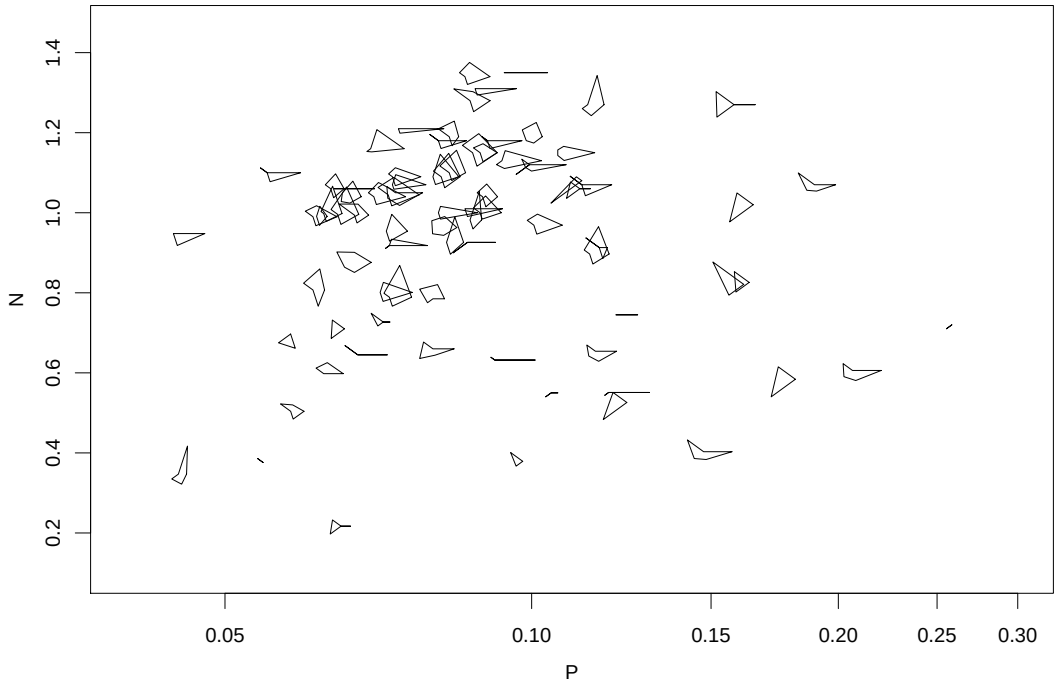

Figure 2.2.a: Scatterplot of Two Soil Variables with Symbols that Show the Number of Specimens of Five Important Plant Species.

b **Symbols.** The symbols that are used in Figur 2.2.a are "stars". The size of the first additional variable is plotted out from the "anchor point" of the observation toward the right, the secod variable in the direction 72^o (for a total of 5 additional variables), the third in the 144^o direction, so top left, etc. The points so obtained are connected to form a five-cornered figure.

There are no limits to creativity for finding such symbols that display the values of additional variables. If only one additional variable should be represented, "bubbles", in other words circles with different radii, are an obvious choice, for two variables rectangles are obvious, and "stars" can be formed in various ways.

"Faces" are an extravagant idea – artificial faces, whose "features" and head shapes correspond to the variable values. The hope is that our eyes can distinguish facial features especially well.

c **Graphical Elements.** Besides these there are still more graphical elements that can encode the data.

- **Flashing** is an especially useful method for a binary variable, but is not usable for reporting.

Further aspects, of which some can reflect quantitative variables and some only nominal (categorial) or binary variables, in order of decreasing effectiveness as judged by the author.

- Size, Color, Orientation (of a line, a rectangle,...), Shape (star, circle,...),
- Intensity or black shading, Hue, Color Saturation, Text (identification number, name of a group,...).

2.3 Dynamic Graphics

a **Dynamic Graphical Elements.** In a script or a book, data can only be represented on the two dimensions of the paper, statically, and we sometimes can't afford color. If data is analyzed on the computer and displayed on a screen, we have a few more possibilities:

- We can "move" representations and thus, for example, achieve a three dimensional impression by looking at rotations of a three dimensional "point cloud."
- We can allow interactions by the user: If a point in the representation is "clicked", the computer can, for example, label this point. Additionally, a representation can be influenced by the usual type of "menu controls".

b **Linked Views.** On the screen, several representations can be shown simulataneously. The simplest variant consists of the already mentioned scatter plot matrix. If in one of these "panels" points are marked interactively, it is very useful if they are simultaneously highlighted in all of the representations. Since the highlighting is usually done with colors, we call this interaction "brushing".

On paper it is hard to write or show anything clever about interactive and dynamic graphics. Try it out!

2.4 Characteristic Numbers

a The distribution of a single variable X is often described by mean value $\bar{x} = \left(\sum_{i=1}^{n} x_i\right)/n$ and standard deviation s. The latter is the root of the (empirical) variance

$$\widehat{\mathrm{var}}\langle X\rangle = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

(The abbreviation var for the variance wears a hat because the "theoretical" variance (without a hat) relates to the model, and $\widehat{\mathrm{var}}$ is an estimator for it.)

b In multivariate statistics we consider multiple variables $X^{(j)}$, $j = 1, 2, ..., m$. Fundamental to the description of the joint distribution is a measurement for the relationship between two variables $X^{(j)}$ and $X^{(k)}$. Analogously to the variance we form the **covariance**

$$\widehat{\text{cov}}\langle X^{(j)}, X^{(k)}\rangle = \frac{1}{n-1}\sum_{i=1}^{n}(x_i^{(j)} - \bar{x}^{(j)})(x_i^{(k)} - \bar{x}^{(k)}) \ .$$

So that the measurement is independent of the measurement units of the variables, we standardize this value by dividing by the standard deviation and get the (product-moment or Pearson) **correlation**

$$\widehat{\rho}\langle X^{(j)}, X^{(k)}\rangle = \frac{\widehat{\text{cov}}\langle X^{(j)}, X^{(k)}\rangle}{\sqrt{\widehat{\text{var}}\langle X^{(j)}\rangle\, \widehat{\text{var}}\langle X^{(k)}\rangle}} \ .$$

It's value lies between -1 and 1. Large values describe a strong "positive" linear relationship; large values of $X^{(1)}$ frequently occur with large values of $X^{(2)}$ and small values behave the same way. Strong negative correlation means that large values of $X^{(1)}$ frequently occur with small values of $X^{(2)}$. Variables that display no (linear) relationship lead to a correlation value near 0.

c Naturally there also exist other ways to measure locations, scatter, and relationship between two variables. Mean, variance, covariance, and Pearson correlation are those that lead to mathematically simple results.

An important reason for considering other characteristic numbers is the lack of **robustness** of the discussed measures: If an outlier occurs, i.e. an observation that "fits with the majority of the observations poorly", it has a large influence on the characteristic numbers, particularly on the variance and covariance. This is often undesirable in the interpretation and inefficient when it comes to the final statistics. Robust methods are therefore important, but we won't treat them in this chapter, except for the following notion.

d **Rank Correlation.** A simple and famous alternative will now be mentioned: As in univariate statistics, we can avoid an excessively large influence of outliers on the correlation by converting data to ranks. For each $X^{(j)}$ we use the **rank transformation** and then calculate the Pearson correlation on the transformed data. This measure is then called **Spearman's rank correlation**. For more detail see Stahel (2002), Ch. 3.3.

However, this idea only partially solves the problem of robustness. If two variables are closely (positively) correlated, then we can introduce an observation with a high value for the first variable and a low value for the second variable, and this outlier will have a large influence on the correlation – also on the rank correlation

2.5 Matrix Notation

a Multivariate statistics need matrix calculations and results from linear algebra. Working without these tools would be like excavating a large pit with a shovel. Here these aids are introduced step by step using their applications to multivariate statistics. The concepts and theorems in sufficient generality are collected in the appendix about linear algebra.

b The data, which is the starting point in multivariate statistics, can be written as a matrix in a natural way,

$$\boldsymbol{x} = \begin{bmatrix} x_1^{(1)} & x_1^{(2)} & \dots & x_1^{(m)} \\ x_2^{(1)} & x_2^{(2)} & \dots & x_2^{(m)} \\ \vdots & \vdots & & \vdots \\ x_n^{(1)} & x_n^{(2)} & \dots & x_n^{(m)} \end{bmatrix}.$$

The element $x_i^{(j)}$ contains the value of the jth variable for the ith observation. Table 2.5.b shows an excerpt from this **data matrix** for the species Iris setosa in the iris flowers example.

We use the first four observations of the first two variables as a concrete **numerical example** for the illustration of the following calculations.

	Sepal Leaves		Petal Leaves	
Nr.	Length	Width	Length	Width
1	5.1	3.5	1.4	0.2
2	4.9	3.0	1.4	0.2
3	4.7	3.2	1.3	0.2
4	4.6	3.1	1.5	0.2
5	5.0	3.6	1.4	0.2
		...		
49	5.3	3.7	1.5	0.2
50	5.0	3.3	1.4	0.2

Table 2.5.b: Data for the Iris Flower Example

c The individual variables correspond to a column of this matrix, or, in other words, a **vector**

$$\underline{x}^{(j)} = \begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \\ \vdots \\ x_n^{(1)} \end{bmatrix}, \qquad \underline{x}^{(2)} = \begin{bmatrix} 3.5 \\ 3.0 \\ 3.2 \\ 3.1 \end{bmatrix}$$

The rows contain the values of the variables for an observation i. Since the vectors

are commonly written vertically, we write

$$\underline{x}_i = \begin{bmatrix} x_i^{(1)} \\ x_i^{(2)} \\ \dots \\ x_i^{(m)} \end{bmatrix}, \qquad \underline{x}_3 = \begin{bmatrix} 4.7 \\ 3.2 \end{bmatrix}$$

If we need to write such a vector as a row, then we form the "row vector" $\underline{x}_i^T = [x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(m)}]$, $\underline{x}_3^T = [4.7,\ 3.2]$ (T for "transposed").

d **Mean.** With the help of the vector

$$\underline{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

sums can be written as a "scalar product" : $\sum_i x_i^{(j)} = \underline{1}^T \underline{x}^{(j)}$. Therefore we have

$$\overline{x}^{(j)} = \tfrac{1}{n}\, \underline{1}^T \underline{x}^{(j)} \ .$$

With the matrix product, we can write the vector of all mean values very simply:

$$\underline{\overline{x}}^T = \tfrac{1}{n} \underline{1}^T \boldsymbol{x} \qquad = \tfrac{1}{4}\,[1,1,1,1] \begin{bmatrix} 5.1 & 3.5 \\ 4.9 & 3.0 \\ 4.7 & 3.2 \\ 4.6 & 3.1 \end{bmatrix} = [4.825,\ 3.2] \ .$$

(The vector $\underline{\overline{x}}$ should again be a column vector. However, the right side of the equation gives a row, so the result is $\underline{\overline{x}}^T$.)

e **Centered Data.** For the variances and covariances we need the "centered variables" $x_i^{(j)} - \overline{x}^{(j)}$. This centered data can also be written in a very short form:

$$\boldsymbol{x}_c = \boldsymbol{x} - \underline{1}\, \underline{\overline{x}}^T \ .$$

We have to note that the last term is an unusual matrix product: While usually a row vector is multiplied with a column vector of equal length (to the scalar product), here the order is switched, which produces a matrix. We see this in an example:

$$\begin{aligned} \boldsymbol{x}_c &= \begin{bmatrix} 5.1 & 3.5 \\ 4.9 & 3.0 \\ 4.7 & 3.2 \\ 4.6 & 3.1 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} [4.825,\ 3.2] = \begin{bmatrix} 5.1 & 3.5 \\ 4.9 & 3.0 \\ 4.7 & 3.2 \\ 4.6 & 3.1 \end{bmatrix} - \begin{bmatrix} 4.825 & 3.2 \\ 4.825 & 3.2 \\ 4.825 & 3.2 \\ 4.825 & 3.2 \end{bmatrix} \\ &= \begin{bmatrix} 0.275 & 0.3 \\ 0.075 & -0.2 \\ -0.125 & 0 \\ -0.225 & -0.1 \end{bmatrix} . \end{aligned}$$

f **Variance-Covariance Matrix.** The expression for the covariance in 2.4.b essentially consists of a sum of products, i.e. a scalar product:

$$\begin{aligned}\widehat{\mathrm{cov}}\langle X^{(j)}, X^{(k)}\rangle &= \tfrac{1}{n-1}\, \underline{x}_c^{(j)T}\, \underline{x}_c^{(k)} \\ \widehat{\mathrm{cov}}\langle X^{(1)}, X^{(2)}\rangle &= \tfrac{1}{n-1}\, [0.275, 0.075, -0.125, -0.225] \begin{bmatrix} 0.3 \\ -0.2 \\ 0 \\ -0.1 \end{bmatrix} = 0.03\ .\end{aligned}$$

We get the variance when, in the general expression, we set $j = k$. We can now write a matrix equation that reflects all variances and covariances:

$$\begin{aligned}\tfrac{1}{n-1}\, \boldsymbol{x}_c^T \boldsymbol{x}_c &= \begin{bmatrix} \widehat{\mathrm{var}}\langle X^{(1)}\rangle & \widehat{\mathrm{cov}}\langle X^{(1)}, X^{(2)}\rangle & \ldots & \widehat{\mathrm{cov}}\langle X^{(1)}, X^{(m)}\rangle \\ \widehat{\mathrm{cov}}\langle X^{(2)}, X^{(1)}\rangle & \widehat{\mathrm{var}}\langle X^{(2)}\rangle & \ldots & \widehat{\mathrm{cov}}\langle X^{(2)}, X^{(m)}\rangle \\ \vdots & \vdots & \ldots & \vdots \\ \widehat{\mathrm{cov}}\langle X^{(m)}, X^{(1)}\rangle & \widehat{\mathrm{cov}}\langle X^{(m)}, X^{(2)}\rangle & \ldots & \widehat{\mathrm{var}}\langle X^{(m)}\rangle \end{bmatrix} \\ &= \widehat{\mathrm{var}}\langle \underline{X}\rangle = \widehat{\boldsymbol{\Sigma}}\ .\end{aligned}$$

The expressions $\widehat{\mathrm{var}}\langle \underline{X}\rangle$ and $\widehat{\boldsymbol{\Sigma}}$ are notations for this matrix, which collects all variances and covariances and is called the (empirical) **variance-covariance matrix** or in short the **covariance matrix**. It has fundamental importance for all of multivariate statistics. The initially somewhat surprising convention to write the variances in the "diagonal" of this matrix will prove itself for mathematical relationships. This is already apparent in the reasoning with which it was introduced here.

In the example

$$\begin{aligned}\widehat{\boldsymbol{\Sigma}} &= \tfrac{1}{n-1} \begin{bmatrix} 0.275 & 0.075 & -0.125 & -0.225 \\ 0.3 & -0.2 & 0 & -0.1 \end{bmatrix} \begin{bmatrix} 0.275 & 0.3 \\ 0.075 & -0.2 \\ -0.125 & 0 \\ -0.225 & -0.1 \end{bmatrix} \\ &= \begin{bmatrix} 0.0492 & 0.0300 \\ 0.0300 & 0.0467 \end{bmatrix}\ .\end{aligned}$$

It is useful to have the covariance matrix be **symmetric**; we have $\widehat{\mathrm{cov}}\langle X^{(j)}, X^{(k)}\rangle = \widehat{\mathrm{cov}}\langle X^{(k)}, X^{(j)}\rangle$.

We will, throughout the script, give hints to the functions of the S language (the R software) that are needed to calculate the methods mentioned. These hints will begin with the "prompt" character of the system, `>`. The covariance matrix is obtained by typing `var`, denoted here by

```
> var
```

g **Correlation Matrix.** The correlations can now be written as

$$\widehat{\rho}\langle X^{(j)}, X^{(k)}\rangle = \widehat{\rho}_{jk} = \frac{\widehat{\mathfrak{Sigma}}_{jk}}{\sqrt{\widehat{\mathfrak{Sigma}}_{jj}\,\widehat{\mathfrak{Sigma}}_{kk}}}$$

and also be collected in a matrix, the **correlation matrix**.

▷ Table 2.5.g shows the bottom half of the correlation matrix. In the "diagonal" ones are often inserted; the correlation of each variable with itself is 1. A graphical illustration of these numbers is given by the scatter plot matrix in 2.1.d. ◁

> `cor`

Sepal.Length	1			
Sepal.Width	0.743	1		
Petal.Length	0.267	0.178	1	
Petal.Width	0.278	0.233	0.332	1
	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width

Table 2.5.g: Correlations for Iris setosa in the Example of the Iris Flowers

h The simplest covariance matrix is the **unit matrix**

$$\boldsymbol{I} = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix}$$

(the matrix with ones in the diagonal and only zeros elsewhere). It means that the components $X^{(j)}$ of $\underline{X}$ have variance 1 (and, if $\bar{x}^{(j)} = 0$, are therefore standardized variables) and that they are uncorrelated.

2.6 Projections and Linear Transformations

a **Linear Combinations of Variables.** In multivariate statistics, linear combinations (translation: weighted sums plus constants) of the variables $X^{(j)}$ play a central roll.

▷ In the **iris flowers example** it can make sense to record the area and the shape of a leaf instead of its length and width. If we first take the logarithm of the measured length and width variables, then the sum of the log values is a good approximation for the log area up to a constant a, which for elliptical leaves amounts to $= \log_{10}\langle\pi/4\rangle = -0.105 \approx -0.1$. The difference of the two log values determines the shape. ◁

We consider the general case of a linear combination $Y = a + b_1X^{(1)} + b_2X^{(2)}$. We thus form the values $y_i = a + b_1x_i^{(1)} + b_2x_i^{(2)}$ and write these with vectors as

$$y_i = a + \underline{b}^T\underline{x}_i \,.$$

For the numerical example, we act as though the given numbers were already the log length and width, so that we can use the previous results. It is then, for example

$$y_3 = -0.1 + [1 \;\; 1] \begin{bmatrix} 4.7 \\ 3.2 \end{bmatrix} = 7.8 \; .$$

b The **mean** of the y_i is

$$\begin{aligned} \bar{y} &= a + \tfrac{1}{n} \sum\nolimits_i \left(b_1 x_i^{(1)} + b_2 x_i^{(2)} \right) = a + \tfrac{1}{n} \left(b_1 \sum\nolimits_i x_i^{(1)} + b_2 \sum\nolimits_i x_i^{(2)} \right) \\ &= a + \left(b_1 \tfrac{1}{n} \sum\nolimits_i x_i^{(1)} + b_2 \tfrac{1}{n} \sum\nolimits_i x_i^{(2)} \right) = a + b_1 \bar{x}^{(1)} + b_2 \bar{x}^{(2)} \\ &= a + \underline{b}^T \underline{\bar{x}} \end{aligned}$$

The formula $\bar{y} = a + \underline{b}^T \underline{\bar{x}}$ still holds if $\underline{x}_i$ includes more than two variables – and the weight vector $\underline{b}$ accordingly has more components.

c The (empirical) **variance** of the y_i can be transformed by a longer calculation with the same goal:

$$\begin{aligned} \widehat{\mathrm{var}}\langle Y \rangle &= \tfrac{1}{n-1} \sum\nolimits_i (y_i - \bar{y})^2 = \tfrac{1}{n-1} \sum\nolimits_i \left(a + b_1 x_i^{(1)} + b_2 x_i^{(2)} - (a + b_1 \bar{x}^{(1)} + b_2 \bar{x}^{(2)}) \right)^2 \\ &= \tfrac{1}{n-1} \sum\nolimits_i \left(b_1 (x_i^{(1)} - \bar{x}^{(1)}) + b_2 (x_i^{(2)} - \bar{x}^{(2)}) \right)^2 \\ &= \tfrac{1}{n-1} \Big(b_1^2 \sum\nolimits_i (x_i^{(1)} - \bar{x}^{(1)})^2 + 2 b_1 b_2 \sum\nolimits_i (x_i^{(1)} - \bar{x}^{(1)})(x_i^{(2)} - \bar{x}^{(2)}) \\ &\qquad + b_2^2 \sum\nolimits_i (x_i^{(2)} - \bar{x}^{(2)})^2 \Big) \\ &= b_1^2 \widehat{\mathrm{var}}\langle X^{(1)} \rangle + 2 b_1 b_2 \widehat{\mathrm{cov}}\langle X^{(1)}, X^{(2)} \rangle + b_2^2 \widehat{\mathrm{var}}\langle X^{(2)} \rangle \; . \end{aligned}$$

This result can now be written elegantly with help of the covariance matrix:

$$\widehat{\mathrm{var}}\langle Y \rangle = [b_1, \, b_2] \begin{bmatrix} \widehat{\mathrm{var}}\langle X^{(1)} \rangle & \widehat{\mathrm{cov}}\langle X^{(1)} X^{(2)} \rangle \\ \widehat{\mathrm{cov}}\langle X^{(1)} X^{(2)} \rangle & \widehat{\mathrm{var}}\langle X^{(2)} \rangle \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \underline{b}^T \widehat{\mathbf{\Sigma}} \underline{b} \; .$$

Again, the last formula also holds for more than two variables $X^{(j)}$.

d **Projection.** The theorem of the cosine from vector geometry describes the close relationship between the scalar product $\underline{b}^T \underline{x}_i$ with the triangle formed by the vectors $\underline{b}$ and $\underline{x}_i$ and the connecting line of their arrowheads (Fig. 2.6.d). We have

$$\underline{b}^T \underline{x}_i = \|\underline{b}\| \, \|\underline{x}_i\| \, \cos\langle \underline{b}, \underline{x}_i \rangle \; ,$$

where $\cos\langle \underline{b}, \underline{x}_i \rangle$ is the cosine of the angle between $\underline{b}$ and $\underline{x}_i$, and $\|\underline{c}\|$ denotes the length of a vector $\underline{c}$. (This length can be written as the square root of the scalar product of $\underline{c}$ with itself, $\|\underline{c}\| = \sqrt{\underline{c}^T \underline{c}}$.)

The Figure also shows that the vector $\widetilde{\underline{x}}_i$, the so-called "projection" of $\underline{x}_i$ onto the direction of $\underline{b}$, has length $\|\underline{x}_i\| \, \cos\langle \underline{b}, \underline{x}_i \rangle = \underline{b}^T \underline{x}_i / \|\underline{b}\|$. It the vector $\underline{b}$ is chosen to have length $\|\underline{b}\| = 1$, then the $y_i = \underline{b}^T \underline{x}_i$ equal the lengths of the projections of the observed vectors $\underline{x}_i$ onto the direction $\underline{b}$.

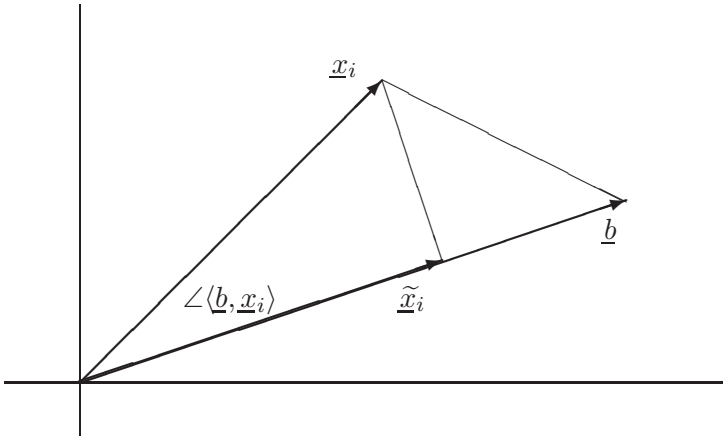

Figure 2.6.d: Display of the Scalar Product, the Angle, and the Projection

If only two variables are present ($m = 2$), then such **unit vectors** or **direction vectors** $\underline{b}$ can be characterized by the angle β which it and the horizontal axis have; they have the form

$$\underline{b} = \begin{bmatrix} \cos\langle\beta\rangle \\ \sin\langle\beta\rangle \end{bmatrix} .$$

In multivariate statistics, projections play an important roll; we will come back to this in the next section.

e **Linear Transformations.** As mentioned above, instead of the (log) length $X^{(1)}$ and width $X^{(2)}$, the Iris flower leaves can also be characterized by the (log) area $Y^{(1)} = a + X^{(1)} + X^{(2)}$ and the shape variable $Y^{(2)} = X^{(2)} - X^{(1)}$. The new values $\underline{Y} = [Y^{(1)}, Y^{(2)}]^T$ come from the old values through a **linear transformation**. In general, we speak of a linear transformation if

$$\underline{y} = \underline{a} + \boldsymbol{B}\underline{x}$$

holds. In the example,

$$\underline{y} = \begin{bmatrix} -0.1 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \underline{x} .$$

f **Important Examples of Linear Transformations.** Transformations can be interpreted and illustrated **geometrically as figures** – especially well in two dimensions. Points $\underline{x}_i$ in the plane go to the image points $\underline{y}_i$.

An especially simple transformation is **stretching** out from the origin. We get the image point $\underline{y}_i$ by multiplying $\underline{x}_i$ componentwise with the stretching factor b. This is a linear transformation with $\underline{a} = \underline{0}$ and

$$\boldsymbol{B} = \begin{bmatrix} b & 0 \\ 0 & b \end{bmatrix} .$$

Matrices that only have numbers different from 0 on the diagonal are called **diagonal matrices** and also written as $\mathrm{diag}\langle b, b\rangle$.

We get a **reflection** about the x-axis with the diagonal matrix $\boldsymbol{B} = \text{diag}\langle 1, -1 \rangle$, and a point reflection about the zero point with $\boldsymbol{B} = \text{diag}\langle -1, -1 \rangle$. With general diagonal matrices we can get stretching and reflections that work in both of the axis directions; $\boldsymbol{B} = \text{diag}\langle b, 1 \rangle$ stretches the data only in the x direction.

A somewhat more unusual representation is shearing. A shearing in the horizontal direction arises with the transformation matrix

$$\boldsymbol{B} = \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix} .$$

Fig. 2.6.f illustrates this representation.

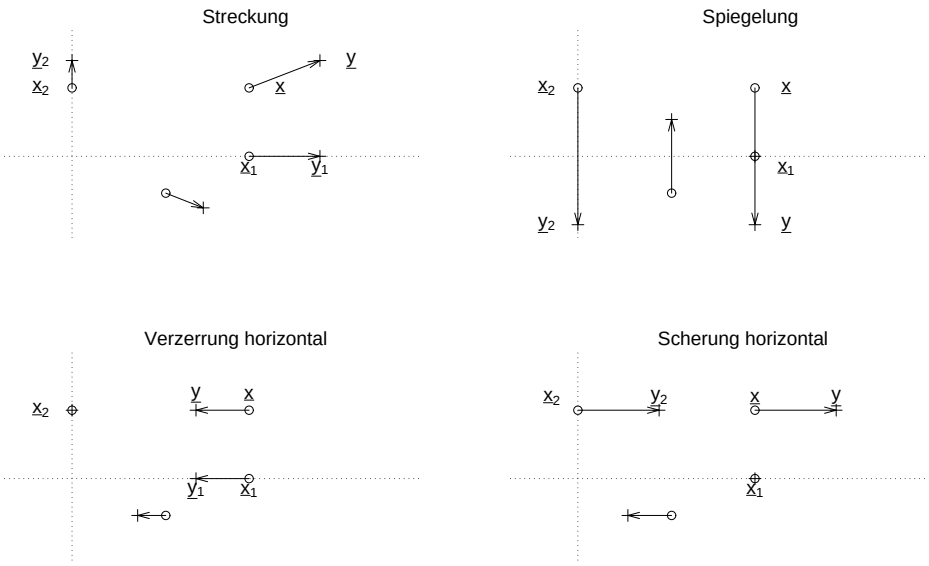

Figure 2.6.f: Four linear representations: Stretching with $b = 1.4$; reflection about the horizontal axis; distortion in the horizontal direction with $b = 0.7$; and shear in the horizontal direction with $a = 0.3$

g **Rotation.** Rotations are an important type of transformation. The corresponding matrices have the form

$$\boldsymbol{B} = \begin{bmatrix} \cos\langle\beta\rangle & -\sin\langle\beta\rangle \\ \sin\langle\beta\rangle & \cos\langle\beta\rangle \end{bmatrix} \quad \text{resp.} \quad = \begin{bmatrix} -\cos\langle\beta\rangle & \sin\langle\beta\rangle \\ \sin\langle\beta\rangle & \cos\langle\beta\rangle \end{bmatrix}$$

How can we see this? Fig. 2.6.g shows that for a point on the horizontal axis with the form $\underline{x}_1 = [x^{(1)} \;\; 0]^T$ the image point

$$\underline{y}_1 = \begin{bmatrix} x^{(1)} \cdot \cos\langle\beta\rangle \\ x^{(1)} \cdot \sin\langle\beta\rangle \end{bmatrix}$$

results, and this is really the same as $\boldsymbol{B}\underline{x}_1$. Similarly, for $\underline{x}_2 = [0, \; x^{(2)}]^T$ we get the image point $\underline{y}_2 = [-x^{(2)} \cdot \sin\langle\beta\rangle \; , \; x^{(2)} \cdot \cos\langle\beta\rangle]^T = \boldsymbol{B}\underline{x}_2$. We get the general point with

the vector $[x^{(1)},\ x^{(2)}]^T$ by summing $\underline{x}_1$ and $\underline{x}_2$. So, we get the corresponding image point by summing $\underline{y}_1$ and $\underline{y}_2$ (as the name "linear" of the transformation expresses and we easily figure out). This gives the result $\underline{y} = \boldsymbol{B}\underline{x}$ with the transformation matrix $\boldsymbol{B}$ written above.

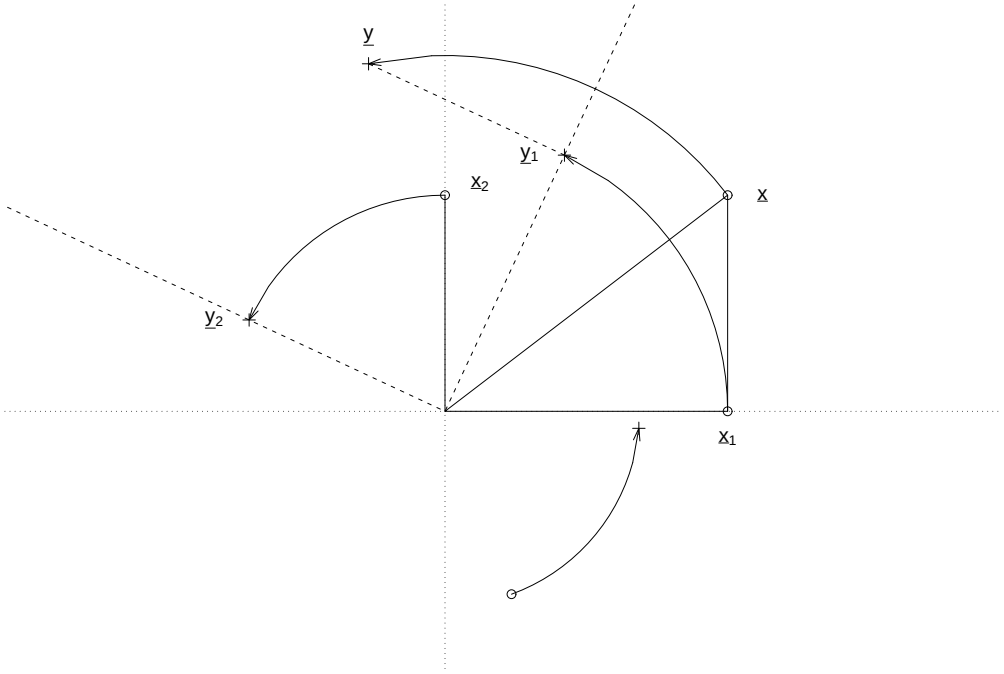

Figure 2.6.g: Rotation by the angle $\beta = 65^o$

In Fig. 2.6.g (ii) the observations in the numerical example (2.5.b) are rotated so that the mean vector lies on the horizontal axis. The rotation angle is $-33.6°$.

h **Two Transformations.** If we carry out two linear transformations one after the other, we can also write the composite tranformation as a single linear transformation, since

$$\begin{aligned} \underline{y} &= \underline{a} + \boldsymbol{B}\underline{x}\ , \quad \widetilde{\underline{y}} = \widetilde{\underline{a}} + \widetilde{\boldsymbol{B}}\underline{y} \\ \widetilde{\underline{y}} &= \widetilde{\underline{a}} + \widetilde{\boldsymbol{B}} \cdot \big(\underline{a} + \boldsymbol{B}\underline{x}\big) = (\widetilde{\underline{a}} + \widetilde{\boldsymbol{B}} \cdot \underline{a}) + (\widetilde{\boldsymbol{B}}\boldsymbol{B})\underline{x} \end{aligned}$$

holds and the result again has the form "vector plus matrix times $\underline{x}$.

i **Mean and Variance of the Transformed Data.** Back to statistics! The mean of the transformed values $\underline{y}_i = \underline{a} + \boldsymbol{B}\underline{x}_i$ is the "composition" of the means of the individual components, for which the result from 2.6.b holds. From this is

$$\overline{\underline{y}} = \underline{a} + \boldsymbol{B}\overline{\underline{x}}\,.$$

The variances of the $Y^{(k)}$ is also obtained from the previous case (2.6.c). And the covariance between the area $Y^{(1)}$ and the shape $Y^{(2)}$? An analogous calculation

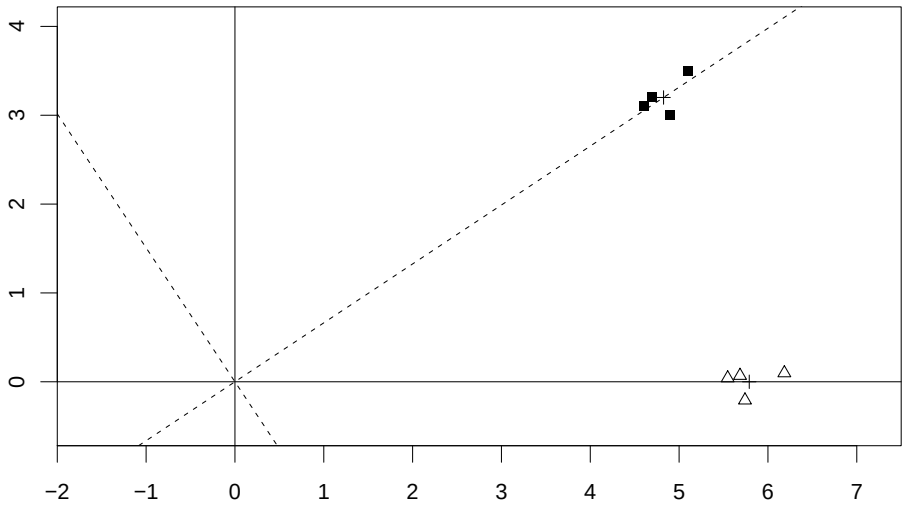

Figure 2.6.g (ii): Rotation of the four observations of the numerical example (2.5.b) by an angle of $-33.6°$. This corresponds to the rotation of the coordinate axes by $+33.6°$. The rotated axes is drawn as dashed.

results if $\underline{b}_1^T$ is the first row of the transformation matrix $\boldsymbol{B}$ and $\underline{b}_2^T$ is the second, $\widehat{\mathrm{cov}}\langle Y^{(1)}, Y^{(2)}\rangle = \underline{b}_1^T \boldsymbol{\Sigma} \underline{b}_2$. By collecting the variances and the covariances, we get

$$\begin{aligned}\widehat{\mathrm{var}}\langle \underline{Y}\rangle &= \boldsymbol{B}\, \widehat{\mathrm{var}}\langle \underline{X}\rangle \boldsymbol{B}^T \\ &= \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 0.0492 & 0.0300 \\ 0.0300 & 0.0467 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 0.0208 & 0.0128 \\ 0.0128 & 0.0751 \end{bmatrix} .\end{aligned}$$

Here $\boldsymbol{B}^T$ is the transposed matrix $\boldsymbol{B}$.

j The derivation of this result proceeds more elegantly if we fall back on the data matrix $\boldsymbol{x}$. The matrix $\boldsymbol{y}$ contains the transformed observation vectors $\underline{y}_i$ as rows $\underline{y}_i^T$. It holds that $\underline{y}_i^T = \underline{a}^T + \underline{x}_i^T \boldsymbol{B}^T$ (since the order of the factors changes under transposition, see 6). Thus is

$$\boldsymbol{y} = \underline{1}\,\underline{a}^T + \boldsymbol{x}\,\boldsymbol{B}^T = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} [a,\ 0] + \begin{bmatrix} 5.1 & 3.5 \\ 4.9 & 3.0 \\ 4.7 & 3.2 \\ 4.6 & 3.1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} .$$

According to 2.5.d the mean is

$$\begin{aligned}\overline{\underline{y}}^T &= \tfrac{1}{n}\underline{1}^T\boldsymbol{y} = \tfrac{1}{n}\underline{1}^T\underline{1}\,\underline{a}^T + \tfrac{1}{n}\underline{1}^T\boldsymbol{x}\boldsymbol{B}^T = \tfrac{1}{n}\,n\,\underline{a}^T + \overline{\underline{x}}^T\boldsymbol{B}^T \\ \overline{\underline{y}} &= \underline{a} + \boldsymbol{B}\,\overline{\underline{x}} = \begin{bmatrix} -0.1 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 4.825 \\ 3.2 \end{bmatrix} = \begin{bmatrix} 7.925 \\ -1.625 \end{bmatrix} .\end{aligned}$$

The centered, transformed data become

$$\boldsymbol{y}_c = \boldsymbol{y} - \underline{1}\,\overline{\underline{y}}^T = \underline{1}\,\underline{a}^T + \boldsymbol{x}\,\boldsymbol{B}^T - \underline{1}\,(\underline{a}^T + \overline{\underline{x}}^T\boldsymbol{B}^T) = (\boldsymbol{x} - \underline{1}\,\overline{\underline{x}}^T)\,\boldsymbol{B}^T = \boldsymbol{x}_c\,\boldsymbol{B}^T .$$

For the covariance matrix we finally get according to 2.5.f

$$\widehat{\mathrm{var}}\langle \underline{Y} \rangle = \tfrac{1}{n-1}\, \boldsymbol{y}_c^T \boldsymbol{y}_c = \tfrac{1}{n-1}\, \boldsymbol{B}\, \boldsymbol{x}_c^T \boldsymbol{x}_c \boldsymbol{B}^T = \boldsymbol{B}\, \widehat{\mathrm{var}}\langle \underline{X} \rangle\, \boldsymbol{B}^T \ .$$

These formulas also hold if $\boldsymbol{B}$ is not square, in other words if from the m X variables, not as many Y variables are produced. If for $\boldsymbol{B}$ we substitute a one-row matrix $\underline{b}^T$, then we get the formula for a linear combination (2.6.b and 2.6.c) as a special case.

k **The Identity.** A question that initially seems rather superfluous: How does the transformation look that changes nothing, for which $\underline{y}_i = \underline{x}_i$ for all possible $\underline{x}_i$? Stupid question? Now, you know that for an understanding of addition of numbers zero is important, which leaves every number unchanged, just like one for multiplication.

With linear transformations we combine an addition of vectors and a multiplication of a vector with a matrix (where the matrix stands to the left of the vector). For the addition of the vectors, the zero vector $\underline{0}$, which consists of only zeros, leaves the other vector unchanged; for the multiplication we need the **unit matrix** $\boldsymbol{I}$:

$$\underline{X} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix} \begin{bmatrix} X^{(1)} \\ X^{(2)} \\ \vdots \\ X^{(m)} \end{bmatrix} = \underline{0} + \boldsymbol{I}\,\underline{X} \ .$$

l **Back Transformation, Inverse Matrix.** In the case of the transformation of (log) length and width to (log) area and shape, we can get the old variables from the new with the reverse transformation

$$\underline{X} = \boldsymbol{B}^{-1}(\underline{Y} - \underline{a}) = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \left(\begin{bmatrix} Y^{(1)} \\ Y^{(2)} \end{bmatrix} - \begin{bmatrix} a \\ 0 \end{bmatrix} \right) \ ,$$

for the example, as we can easily calculate. The matrix $\boldsymbol{B}^{-1}$ is the **inverse** of the matrix $\boldsymbol{B}$. It holds that $\boldsymbol{B}^{-1}\boldsymbol{B} = \boldsymbol{I}$, since the transformation with the matrix $\boldsymbol{B}$, followed by the reverse transformation, gives the identity. Such a matrix can exist only for square matrices, but not all square matrices have an inverse. Those that have an inverse are called **regular**, **invertible** or **nonsingular** matrices, while the others are called **singular** (see 2.A.0.j in the appendix).

m **Standardization.** In univariate statistics it is often useful to standardize a sample X by removing the mean and dividing by the standard deviation; we form $z_i = (x_i - \bar{x})/\widehat{\sigma}$. From the data $\boldsymbol{x}$ with mean vector $\underline{\overline{x}}$ and empirical covariance matrix $\widehat{\boldsymbol{\Sigma}}$ we now want to get a data matrix $\boldsymbol{z}$ with $\underline{\overline{z}} = \underline{0}$ and $\widehat{\mathrm{var}}\langle \underline{Z} \rangle = \boldsymbol{I}$ through linear transformation.

It is easy to get $\underline{\overline{z}} = \underline{0}$: We only need to remove the mean vector $\underline{\overline{x}}$ from each $\underline{x}_i$.

For the second condition we need a result from linear algebra: The covariance matrix $\widehat{\boldsymbol{\Sigma}}$ is symmetric and called **positive semi-definite**, which means that for any given vector $\underline{b}$ (that is not $= \underline{0}$),

$$\underline{b}^T \widehat{\boldsymbol{\Sigma}}\, \underline{b} \geq 0 \ .$$

For each such matrix, according to a result from linear algebra, a matrix $\boldsymbol{B}$ can be found, so that $\boldsymbol{B}\,\boldsymbol{B}^T = \widehat{\boldsymbol{\Sigma}}$ – there are infinitely many. The so-called **Cholesky**

Decomposition gives one of these in the form of a triangle matrix – in the example

$$\boldsymbol{B} = \begin{bmatrix} 0.222 & 0 \\ 0.135 & 0.168 \end{bmatrix}, \quad \begin{bmatrix} 0.222 & 0 \\ 0.135 & 0.168 \end{bmatrix} \begin{bmatrix} 0.222 & 0.135 \\ 0 & 0.168 \end{bmatrix} = \begin{bmatrix} 0.0492 & 0.0300 \\ 0.0300 & 0.0467 \end{bmatrix}.$$

Now we take transformation $\underline{z} = \boldsymbol{C}(\underline{x} - \widehat{\underline{\mu}})$, with $\boldsymbol{C} = \boldsymbol{B}^{-1}$, and look at it! We have

$$\begin{aligned} \underline{z}_i &= \boldsymbol{C}(\underline{x}_i - \overline{\underline{x}}) = -\boldsymbol{C}\overline{\underline{x}} + \boldsymbol{C}\underline{x}_i \\ \overline{\underline{z}} &= \boldsymbol{C}(\overline{\underline{x}} - \overline{\underline{x}}) = \underline{0} \\ \widehat{\mathrm{var}}\langle \underline{Z} \rangle &= \boldsymbol{C}\widehat{\boldsymbol{\Sigma}}\boldsymbol{C}^T = \boldsymbol{C}\boldsymbol{B}\boldsymbol{B}^T\boldsymbol{C}^T = \boldsymbol{C}\boldsymbol{C}^{-1}(\boldsymbol{C}^{-1})^T\boldsymbol{C}^T = \boldsymbol{I}\,. \end{aligned}$$

Thus, the goal is achieved! (However, it is implicitly assumed that $\boldsymbol{C}$ is invertible, so it should be nonsingular, which is equivalent to the condition that $\widehat{\boldsymbol{\Sigma}}$ is nonsingular.) In the example

$$\boldsymbol{C} = \begin{bmatrix} 4.51 & 0 \\ -3.62 & 5.94 \end{bmatrix}, \qquad \boldsymbol{z} = \boldsymbol{x}_c\boldsymbol{C}^T = \begin{bmatrix} 1.240 & 0.785 \\ 0.338 & -1.458 \\ -0.564 & 0.452 \\ -1.014 & 0.221 \end{bmatrix}$$

* As mentioned, the Cholesky decomposition does not give the only matrix $\boldsymbol{B}$ for which $\boldsymbol{B}\,\boldsymbol{B}^T = \widehat{\boldsymbol{\Sigma}}$ holds. We will come back to other solutions.

n **Back Transformation of a Rotation, Orthogonal Matrices.** To reverse a rotation by the angle β, we have to rotate again, this time by $-\beta$. This leads to the matrix

$$\begin{bmatrix} \cos\langle -\beta \rangle & -\sin\langle -\beta \rangle \\ \sin\langle -\beta \rangle & \cos\langle -\beta \rangle \end{bmatrix} = \begin{bmatrix} \cos\langle \beta \rangle & \sin\langle \beta \rangle \\ -\sin\langle \beta \rangle & \cos\langle \beta \rangle \end{bmatrix} = \boldsymbol{B}^T\,.$$

So, the transposed matrix $\boldsymbol{B}^T$ is the same as the inverse, $\boldsymbol{B}^{-1}$, and thus

$$\boldsymbol{B}^T\boldsymbol{B} = \boldsymbol{I}\,.$$

This also holds for reflections. Matrices with this property are called **orthogonal** matrices, and the corresponding transformations are also called "orthogonal" .

The central property of orthogonal transformations is that they do not change the distance between points and thus all "shapes" or patterns of points remain. It can easily be seen that the (squared) lengths of vectors remain unchanged by orthogonal transformations:

$$\|\underline{y}_i\|^2 = \underline{y}_i^T\underline{y}_i = \underline{x}_i^T\boldsymbol{B}^T\boldsymbol{B}\underline{x}_i = \underline{x}_i^T\boldsymbol{I}\underline{x}_i = \underline{x}_i^T\underline{x}_i = \|\underline{x}_i\|^2\,.$$

Note. Here we have implicitly assumed that no shift takes place, so $\underline{a} = \underline{0}$ and the rotation occurs around zero. We can also consider more general rotations, with a shift; then only the lengths of differences $\underline{x}_i - \underline{x}_h$ are preserved.

o* We come back to the question of more solutions of $\boldsymbol{B}\boldsymbol{B}^T = \widehat{\boldsymbol{\Sigma}}$! If $\boldsymbol{B}_c$ is a solution, for example the Cholesky root, then also $\boldsymbol{B}_c\boldsymbol{B}_o$ is a solution with every orthogonal matrix $\boldsymbol{B}_o$, since $\boldsymbol{B}_c\boldsymbol{B}_o(\boldsymbol{B}_c\boldsymbol{B}_o)^T = \boldsymbol{B}_c\boldsymbol{B}_o\boldsymbol{B}_o^T\boldsymbol{B}_c^T = \boldsymbol{B}_c\boldsymbol{I}\,\boldsymbol{B}_c^T = \boldsymbol{B}_c\boldsymbol{B}_c^T = \widehat{\boldsymbol{\Sigma}}$. We can also prove the reverse in a similarly simple way: Two solutions always differ by an orthogonal "factor" $\boldsymbol{B}_o$.

We can easily make it clear what is behind this: If we have standardized data and transform it with an orthogonal matrix, i.e. rotate or reflecte it, then the data remains standardized. So, if we compose a "standardization transformation" and an orthogonal transformation, we again have a standardization transformation.

p **Basis Transformation.** Instead of thinking about the rotation of all points by the angle β we can also talk about a rotation of the coordinate system by $-\beta$ – both concepts lead to the same "new coordinates" $\underline{y}_i$. In Figure 2.6.g (i) this new coordinate system is shown dashed. Even in more than two dimensions, orthogonal transformations are equivalent to changing the coordinate system. The length of vectors remain the same under this change.

2.7 Projection Pursuit

a **Basic Idea.** Explorative multivariate statistics should find interesting structures in the data. If these don't show up in a scatter plot matrix, we can hope that they become apparent under appropriate transformation of the coordinate system. We thus seek "directions in space" that show interesting structures.

b **Manual Search.** Dynamic graphics programs allow the two axes that are used for a scatter plot to be changed continuously and as desired with help of some kind of movement of a joystick, the cursors, or similar inputs (cf. 2.3.a). Then structures can be searched for by eye.

What is shown on the screen is a scatter plot of two projections $\underline{y}^{(1)} = \boldsymbol{x}\underline{b}_1$ and $\underline{y}^{(2)} = \boldsymbol{x}\underline{b}_2$ of the data (2.6.d) with two perpendicular direction vectors $\underline{b}_1$ and $\underline{b}_2$. Expressed another way, we use the first two coordinates of the points in a new coordinate system with orthogonal transformation matrix (2.6.p).

c Interesting projections can also be sought by the computer with numerical optimization if we can give a quantitative measure for the "interestingness" of a projection, a **projection index** Q. This measure depends on the application.

d* Which properties should a projection index have? Usually, deviations from the normal distribution are interesting. Since such a deviation remains the same if we alter the scale (location and scatter) of the data, such a Q should be "**invariant**" to shifts and stretching:

$$Q\langle a + b\underline{y}\rangle = Q\langle\underline{y}\rangle$$

We therefore define Q for standardized values and use $Q\langle\underline{y}\rangle = Q_0\langle(\underline{y} - \bar{y})/s_y\rangle$ where $s_y^2 = \widehat{\mathrm{var}}\langle Y\rangle$ is the empirical variance of $\underline{y}$. For calculation we standardize $\boldsymbol{X}$ (see 2.6.m). Then $s_y = 1$ for all direction vectors $\underline{b}$ with length $|\underline{v}| = 1$.

e **Projection Indices.** Two classes are common:

1. Functionals for densities. We first estimate the density with a d dimensional density estimator.
2. Higher empirical moments (third and fourth)

2.S S-Functions

a **Graphics Functions:** The most important graphing functions for representation of multiple variables are `pairs, symbols, coplot`.

b **Function** `pairs`**.** creates a scatter plot matrix of the data. Figur 2.1.a comes from

> `pairs(iris[,1:4], pch=c(0,1,2)[iris[,5]])`

With the arguments `diag.panel, lower.panel` and `upper.panel` the diagonals and the upper or lower triangle matrix can be changed. We can thus, for example, put the pairwise correlations of the variables in the upper triangle matrix, which would otherwise just contain the mirrored scatter plots from below.

```
> data(iris)
> pairs(iris, lower.panel=panel.smooth,
    upper.panel= function(x,y)
      text(mean(range(x)),mean(range(y)), round(cor(x,y),3) ) )
```

For the lower triangle, the useful function `panel.smooth` is applied here. (In the example, this is of limited use, and is meaningless for the bottom line.)
`example(pairs)` shows how we get histograms for the diagonal.

c **Function** `symbols`**.** Scatter diagram with symbols.
In addition to the two variables that are represented as horizontal and vertical coordinates in the scatter diagram, other variables are represented by the use of various types of symbols.

```
Usage:
  symbols(x, y = NULL, circles, squares, rectangles, stars,
          thermometers, boxplots, inches = TRUE, add = FALSE,
          fg = 1, bg = NA, xlab = NULL, ylab = NULL, main = NULL,
          xlim = NULL, ylim = NULL, ...)
```

Of the arguments `circles, squares, rectangles, stars, thermometers` and `boxplots` only one is used.

- With `circles=z` circles with a radius proportional to `z` are shown. (`z` is a vector with an elements for each observation.) `squares=z` graphs corresponding squares.
- The argument `rectangles=cbind(z1,z2)`, generates rectangles where the side lengths reflect the two variables `z1` and `z2`.
- The argument `thermometers` can take a matrix with 3 or 4 columns that determine the width, height, and filling of the "thermometer" . If we only have 3

variables, we want to show the third as the height of the filling of the thermometer. We must then give the argument `thermometers` a matrix in which the first two columns are filled with ones `thermometers=cbind(1,1,z)`

- The argument `stars=mat` with a matrix `mat` with m columns draws polygons (m corners). Each direction from the "center " of a "star " to a corner symbolizes an additional variable. The distance from the center to the corner gives the relative size of this variable.

The function `stars` is however much more flexible and extensive for making this type of plot, see `par(ask=TRUE); example(stars)`.

Via colors and line type, more variables can be displayed. We can thus generate very dense – even overloaded – graphical representations, as long as there are not too many observations.

d **Function `coplot`.** : Matrices of "conditional scatter diagrams".

```
> coplot(lat ~ long | depth * mag, data = quakes)
```

separates the observations of the data set `quakes` into subgroups (that overlap) corresponding to the variables `depth` and `mag` and shows for each subgroup a scatter diagram of the variables `lat` against `long`. The first argument is a `formula` that consists of the variables from the dataset, which is designated with the argument `data` (or possible other vectors). Left of the conditioning indicators | are the two variables that are used for drawing the scatter diagram, on the right the conditions variables.

e **Dynamic Graphics and Projection Pursuit.** The two packages `xgobi` and Rggobi form a bridge to two versions of a corresponding program for dynamic graphics, that also contains projection pursuit methods.

f **Characteristic Numbers**
`t.x <- as.matrix(iris[1:50,1:4])` : Conversion into a matrix, only first plant type (observations 1 to 50)
`(t.mn <- apply(t.x, 2, mean))` : Mean values for the columns
`(t.var <- var(t.x))` : Variance-Covariance Matrix

g **Standardization**
`t.xc <- scale(t.x, scale=FALSE)` : Centered Observations.
`scale(t.x)` with default value `scale=TRUE` standardizes the individual variables. `(t.bt <- t(chol(t.var)))` : "factorization of the covariance matrix".
`t.bt%*%t(t.bt)` reflects the covariance matrix.
`(t.b <- solve(t.bt))` : Inversion of the matrix $\widetilde{B}$.
`t.z <- t.xc %*% t(t.b)` : Standardization.
Check of the results with `apply(t.z,2,mean)` and `var(t.z)`

h **Distribution of the Lengths of the Standardized Observations**
`t.d2 <- apply(t.z^2,1,sum)`
Quantile-Quantile Diagram:

```
qqplot(qchisq(ppoints(length(t.d2)),ncol(t.z)),t.d2,
  xlab="Quantiles of the Chisq. Distr.", ylab="Ordered Mahalanobis Dist.",
  main="QQ-plot for Mahalanobis Distances")
```