Chapter 10
Principal Components Analysis

Chapter 9 presented the basic geometric tools needed to produce a lower dimensional description of the rows and columns of a multivariate data matrix. Principal components analysis has the same objective with the exception that the rows of the data matrix $\mathcal{X}$ will now be considered as observations from a p-variate random variable X. The principle idea of reducing the dimension of X is achieved through linear combinations. Low dimensional linear combinations are often easier to interpret and serve as an intermediate step in a more complex data analysis. More precisely one looks for linear combinations which create the largest spread among the values of X. In other words, one is searching for linear combinations with the largest variances.

Section 10.1 introduces the basic ideas and technical elements behind principal components. No particular assumption will be made on X except that the mean vector and the covariance matrix exist. When reference is made to a data matrix $\mathcal{X}$ in Section 10.2, the empirical mean and covariance matrix will be used. Section 10.3 shows how to interpret the principal components by studying their correlations with the original components of X. Often analyses are performed in practice by looking at two-dimensional scatterplots. Section 10.4 develops inference techniques on principal components. This is particularly helpful in establishing the appropriate dimension reduction and thus in determining the quality of the resulting lower dimensional representations. Since principal component analysis is performed on covariance matrices, it is not scale invariant. Often, the measurement units of the components of X are quite different, so it is reasonable to standardize the measurement units. The normalized version of principal components is defined in Section 10.5. In Section 10.6 it is discovered that the empirical principal components are the factors of appropriate transformations of the data matrix. The classical way of defining principal components through linear combinations with respect to the largest variance is described here in geometric terms, i.e., in terms of the optimal fit within subspaces generated by the columns and/or the rows of $\mathcal{X}$ as was discussed in Chapter 9. Section 10.9 concludes with additional examples.

W.K. Härdle, L. Simar, *Applied Multivariate Statistical Analysis*,
DOI 10.1007/978-3-642-17229-8_10,

10.1 Standardized Linear Combination

The main objective of principal components analysis (PC) is to reduce the dimension of the observations. The simplest way of dimension reduction is to take just one element of the observed vector and to discard all others. This is not a very reasonable approach, as we have seen in the earlier chapters, since strength may be lost in interpreting the data. In the bank notes example we have seen that just one variable (e.g. $X_1 = \text{length}$) had no discriminatory power in distinguishing counterfeit from genuine bank notes. An alternative method is to weight all variables equally, i.e., to consider the simple average $p^{-1}\sum_{j=1}^{p} X_j$ of all the elements in the vector $X = (X_1, \dots, X_p)^\top$. This again is undesirable, since all of the elements of X are considered with equal importance (weight).

A more flexible approach is to study a weighted average, namely

$$\delta^\top X = \sum_{j=1}^{p} \delta_j X_j, \qquad \text{such that} \quad \sum_{j=1}^{p} \delta_j^2 = 1. \tag{10.1}$$

The weighting vector $\delta = (\delta_1, \dots, \delta_p)^\top$ can then be optimized to investigate and to detect specific features. We call (10.1) a standardized linear combination (SLC). Which SLC should we choose? One aim is to maximize the variance of the projection $\delta^\top X$, i.e., to choose δ according to

$$\max_{\{\delta:\|\delta\|=1\}} \mathsf{Var}(\delta^\top X) = \max_{\{\delta:\|\delta\|=1\}} \delta^\top \mathsf{Var}(X)\delta. \tag{10.2}$$

The interesting "directions" of δ are found through the spectral decomposition of the covariance matrix. Indeed, from Theorem 2.5, the direction δ is given by the eigenvector γ_1 corresponding to the largest eigenvalue λ_1 of the covariance matrix $\Sigma = \mathsf{Var}(X)$.

Figures 10.1 and 10.2 show two such projections (SLCs) of the same data set with zero mean. In Figure 10.1 an arbitrary projection is displayed. The upper window shows the data point cloud and the line onto which the data are projected. The middle window shows the projected values in the selected direction. The lower window shows the variance of the actual projection and the percentage of the total variance that is explained.

Figure 10.2 shows the projection that captures the majority of the variance in the data. This direction is of interest and is located along the main direction of the point cloud. The same line of thought can be applied to all data orthogonal to this direction leading to the second eigenvector. The SLC with the highest variance obtained from maximizing (10.2) is the first principal component (PC) $y_1 = \gamma_1^\top X$. Orthogonal to the direction γ_1 we find the SLC with the second highest variance: $y_2 = \gamma_2^\top X$, the second PC.

Proceeding in this way and writing in matrix notation, the result for a random variable X with $\mathsf{E}(X) = \mu$ and $\mathsf{Var}(X) = \Sigma = \Gamma\Lambda\Gamma^\top$ is the PC transformation which is defined as

$$Y = \Gamma^\top (X - \mu). \tag{10.3}$$

Here we have centered the variable X in order to obtain a zero mean PC variable Y.

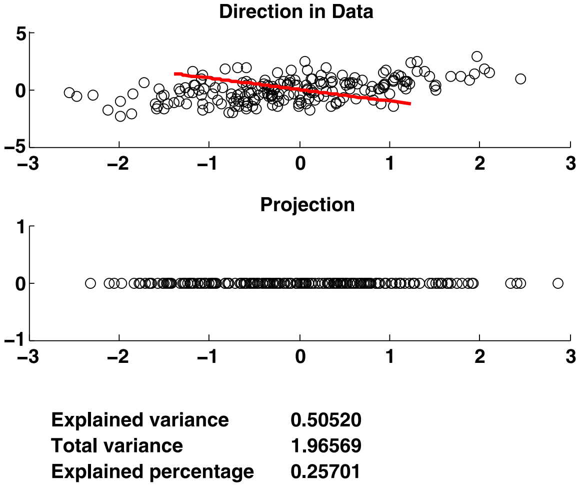


Fig. 10.1 An arbitrary SLC `MVApcasimu`

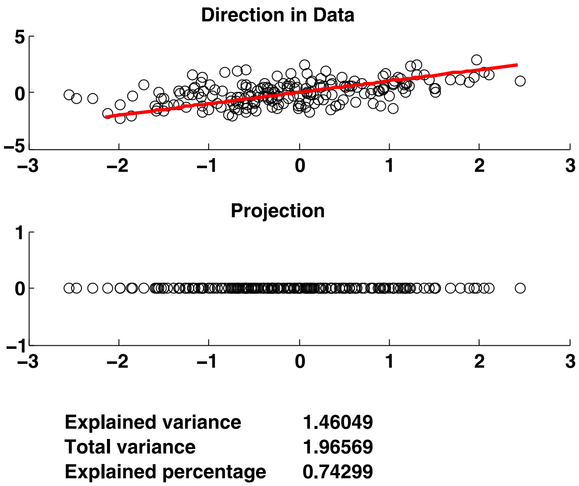


Fig. 10.2 The most interesting SLC `MVApcasimu`

Example 10.1 Consider a bivariate normal distribution $N(0, \Sigma)$ with $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ and $\rho > 0$ (see Example 3.13). Recall that the eigenvalues of this matrix are $\lambda_1 = 1 + \rho$ and $\lambda_2 = 1 - \rho$ with corresponding eigenvectors

$$\gamma_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \gamma_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

The PC transformation is thus

$$Y = \Gamma^\top (X - \mu) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} X$$

or

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} X_1 + X_2 \\ X_1 - X_2 \end{pmatrix}.$$

So the first principal component is

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

and the second is

$$Y_2 = \frac{1}{\sqrt{2}}(X_1 - X_2).$$

Let us compute the variances of these PCs using formulas (4.22)–(4.26):

$$\begin{aligned}
\mathsf{Var}(Y_1) = \mathsf{Var}\left\{\frac{1}{\sqrt{2}}(X_1 + X_2)\right\} &= \frac{1}{2}\mathsf{Var}(X_1 + X_2) \\
&= \frac{1}{2}\{\mathsf{Var}(X_1) + \mathsf{Var}(X_2) + 2\,\mathsf{Cov}(X_1, X_2)\} \\
&= \frac{1}{2}(1 + 1 + 2\rho) = 1 + \rho \\
&= \lambda_1.
\end{aligned}$$

Similarly we find that

$$\mathsf{Var}(Y_2) = \lambda_2.$$

This can be expressed more generally and is given in the next theorem.

Theorem 10.1 *For a given $X \sim (\mu, \Sigma)$ let $Y = \Gamma^\top(X - \mu)$ be the PC transformation. Then*

$$\mathsf{E}\,Y_j = 0, \quad j = 1, \ldots, p \tag{10.4}$$

$$\mathsf{Var}(Y_j) = \lambda_j, \quad j = 1, \ldots, p \tag{10.5}$$

$$\mathsf{Cov}(Y_i, Y_j) = 0, \quad i \neq j \tag{10.6}$$

$$\mathsf{Var}(Y_1) \geq \mathsf{Var}(Y_2) \geq \cdots \geq \mathsf{Var}(Y_p) \geq 0 \tag{10.7}$$

$$\sum_{j=1}^{p} \mathsf{Var}(Y_j) = \mathrm{tr}(\Sigma) \tag{10.8}$$

$$\prod_{j=1}^{p} \mathsf{Var}(Y_j) = |\Sigma|. \tag{10.9}$$

Proof To prove (10.6), we use γ_i to denote the ith column of Γ. Then

$$\mathsf{Cov}(Y_i, Y_j) = \gamma_i^\top \mathsf{Var}(X - \mu)\gamma_j = \gamma_i^\top \mathsf{Var}(X)\gamma_j.$$

As $\mathsf{Var}(X) = \Sigma = \Gamma\Lambda\Gamma^\top$, $\Gamma^\top\Gamma = \mathcal{I}$, we obtain via the orthogonality of Γ:

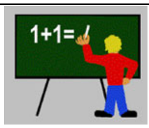
$$\gamma_i^\top \Gamma\Lambda\Gamma^\top \gamma_j = \begin{cases} 0 & i \neq j \\ \lambda_i & i = j. \end{cases}$$

In fact, as $Y_i = \gamma_i^\top(X - \mu)$ lies in the eigenvector space corresponding to γ_i, and eigenvector spaces corresponding to different eigenvalues are orthogonal to each other, we can directly see Y_i and Y_j are orthogonal to each other, so their covariance is 0. □

The connection between the PC transformation and the search for the best SLC is made in the following theorem, which follows directly from (10.2) and Theorem 2.5.

Theorem 10.2 *There exists no SLC that has larger variance than* $\lambda_1 = \mathsf{Var}(Y_1)$.

Theorem 10.3 *If* $Y = a^\top X$ *is a SLC that is not correlated with the first* k *PCs of* X, *then the variance of* Y *is maximized by choosing it to be the* $(k+1)$*-st PC.*



Summary

↪ A standardized linear combination (SLC) is a weighted average $\delta^\top X = \sum_{j=1}^{p} \delta_j X_j$ where δ is a vector of length 1.

↪ Maximizing the variance of $\delta^\top X$ leads to the choice $\delta = \gamma_1$, the eigenvector corresponding to the largest eigenvalue λ_1 of $\Sigma = \mathsf{Var}(X)$.
This is a projection of X into the one-dimensional space, where the components of X are weighted by the elements of γ_1. $Y_1 = \gamma_1^\top(X - \mu)$ is called the first principal component (PC).

↪ This projection can be generalized for higher dimensions. The PC transformation is the linear transformation $Y = \Gamma^\top(X - \mu)$, where $\Sigma = \mathsf{Var}(X) = \Gamma\Lambda\Gamma^\top$ and $\mu = \mathsf{E}\,X$.
$Y_1, Y_2, \dots, Y_p$ are called the first, second, …, and p-th PCs.

↪ The PCs have zero means, variance $\mathsf{Var}(Y_j) = \lambda_j$, and zero covariances. From $\lambda_1 \geq \cdots \geq \lambda_p$ it follows that $\mathsf{Var}(Y_1) \geq \cdots \geq \mathsf{Var}(Y_p)$. It holds that $\sum_{j=1}^{p} \mathsf{Var}(Y_j) = \operatorname{tr}(\Sigma)$ and $\prod_{j=1}^{p} \mathsf{Var}(Y_j) = |\Sigma|$.

↪ If $Y = a^\top X$ is a SLC which is not correlated with the first k PCs of X then the variance of Y is maximized by choosing it to be the $(k+1)$-st PC.

10.2 Principal Components in Practice

In practice the PC transformation has to be replaced by the respective estimators: μ becomes $\overline{x}$, Σ is replaced by $\mathcal{S}$, etc. If g_1 denotes the first eigenvector of $\mathcal{S}$, the first principal component is given by $y_1 = (\mathcal{X} - 1_n\overline{x}^\top)g_1$. More generally if $\mathcal{S} = \mathcal{G}\mathcal{L}\mathcal{G}^\top$ is the spectral decomposition of $\mathcal{S}$, then the PCs are obtained by

$$\mathcal{Y} = (\mathcal{X} - 1_n\overline{x}^\top)\mathcal{G}. \tag{10.10}$$

Note that with the centering matrix $\mathcal{H} = \mathcal{I} - (n^{-1}1_n 1_n^\top)$ and $\mathcal{H}1_n\overline{x}^\top = 0$ we can write

$$\begin{aligned} \mathcal{S}_{\mathcal{Y}} &= n^{-1}\mathcal{Y}^\top\mathcal{H}\mathcal{Y} = n^{-1}\mathcal{G}^\top(\mathcal{X} - 1_n\overline{x}^\top)^\top\mathcal{H}(\mathcal{X} - 1_n\overline{x}^\top)\mathcal{G} \\ &= n^{-1}\mathcal{G}^\top\mathcal{X}^\top\mathcal{H}\mathcal{X}\mathcal{G} = \mathcal{G}^\top\mathcal{S}\mathcal{G} = \mathcal{L} \end{aligned} \tag{10.11}$$

where $\mathcal{L} = \operatorname{diag}(\ell_1, \ldots, \ell_p)$ is the matrix of eigenvalues of $\mathcal{S}$. Hence the variance of y_i equals the eigenvalue ℓ_i!

The PC technique is sensitive to scale changes. If we multiply one variable by a scalar we obtain different eigenvalues and eigenvectors. This is due to the fact that an eigenvalue decomposition is performed on the covariance matrix and not on the correlation matrix (see Section 10.5). The following warning is therefore important:

! The PC transformation should be applied to data that have approximately the same scale in each variable.

Example 10.2 Let us apply this technique to the bank data set. In this example we do not standardize the data. Figure 10.3 shows some PC plots of the bank data set. The genuine and counterfeit bank notes are marked by "o" and "+" respectively.

Recall that the mean vector of $\mathcal{X}$ is

$$\overline{x} = (214.9,\ 130.1,\ 129.9,\ 9.4,\ 10.6,\ 140.5)^\top.$$

The vector of eigenvalues of $\mathcal{S}$ is

$$\ell = (2.985,\ 0.931,\ 0.242,\ 0.194,\ 0.085,\ 0.035)^\top.$$

The eigenvectors g_j are given by the columns of the matrix

$$\mathcal{G} = \begin{pmatrix} -0.044 & 0.011 & 0.326 & 0.562 & -0.753 & 0.098 \\ 0.112 & 0.071 & 0.259 & 0.455 & 0.347 & -0.767 \\ 0.139 & 0.066 & 0.345 & 0.415 & 0.535 & 0.632 \\ 0.768 & -0.563 & 0.218 & -0.186 & -0.100 & -0.022 \\ 0.202 & 0.659 & 0.557 & -0.451 & -0.102 & -0.035 \\ -0.579 & -0.489 & 0.592 & -0.258 & 0.085 & -0.046 \end{pmatrix}.$$

The first column of $\mathcal{G}$ is the first eigenvector and gives the weights used in the linear combination of the original data in the first PC.

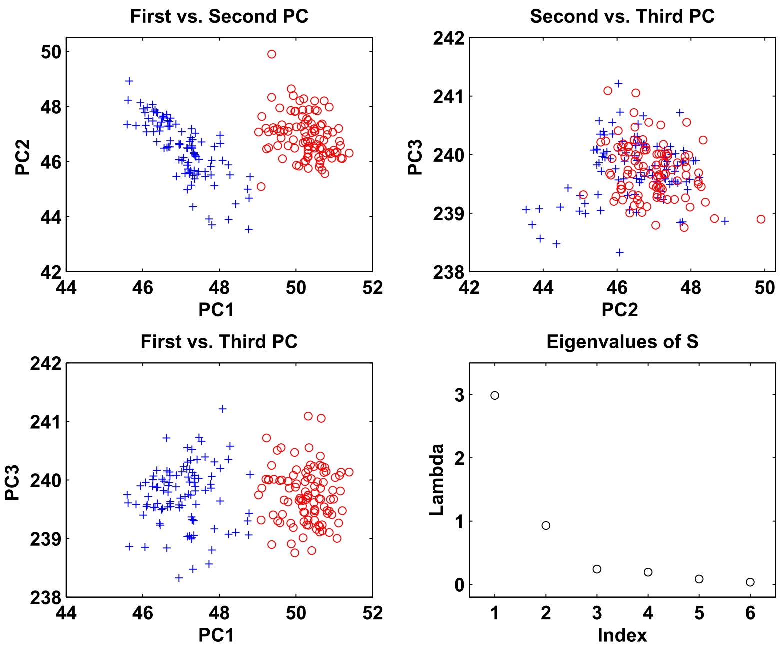


Fig. 10.3 Principal components of the bank data MVApcabank

Example 10.3 To see how sensitive the PCs are to a change in the scale of the variables, assume that X_1, X_2, X_3 and X_6 are measured in *cm* and that X_4 and X_5 remain in *mm* in the bank data set. This leads to:

$$\bar{x} = (21.49,\ 13.01,\ 12.99,\ 9.41,\ 10.65,\ 14.05)^{\top}.$$

The covariance matrix can be obtained from S in (3.4) by dividing rows 1, 2, 3, 6 and columns 1, 2, 3, 6 by 10. We obtain:

$$\ell = (2.101,\ 0.623,\ 0.005,\ 0.002,\ 0.001,\ 0.0004)^{\top}$$

which clearly differs from Example 10.2. Only the first two eigenvectors are given:

$$g_1 = (-0.005,\ 0.011,\ 0.014,\ 0.992,\ 0.113,\ -0.052)^{\top}$$
$$g_2 = (-0.001,\ 0.013,\ 0.016,\ -0.117,\ 0.991,\ -0.069)^{\top}.$$

Comparing these results to the first two columns of $\mathcal{G}$ from Example 10.2, a completely different story is revealed. Here the first component is dominated by X_4 (lower margin) and the second by X_5 (upper margin), while all of the other variables have much less weight. The results are shown in Figure 10.4. Section 10.5 will show how to select a reasonable standardization of the variables when the scales are too different.

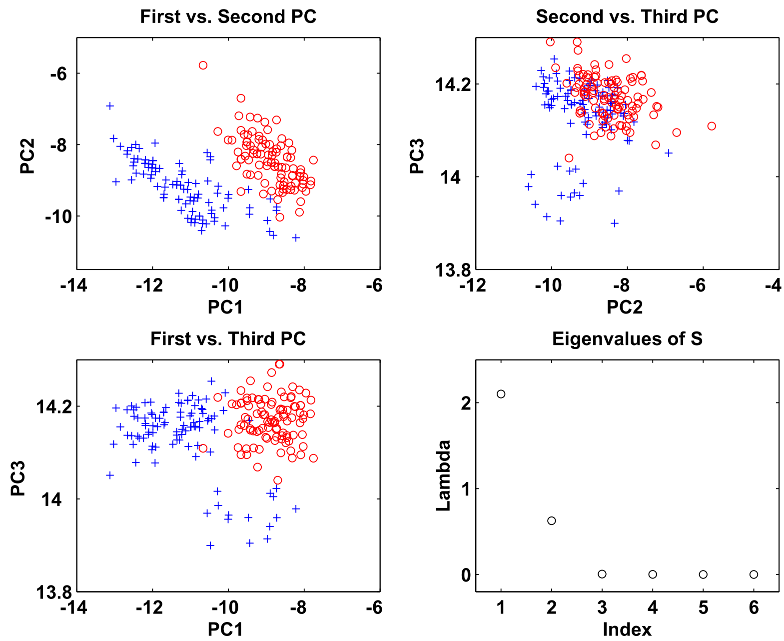
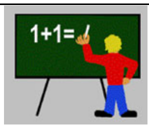


Fig. 10.4 Principal components of the rescaled bank data `MVApcabankr`



Summary

↪ The scale of the variables should be roughly the same for PC transformations.

↪ For the practical implementation of principal components analysis (PCA) we replace μ by the mean $\overline{x}$ and Σ by the empirical covariance $\mathcal{S}$. Then we compute the eigenvalues $\ell_1, \dots, \ell_p$ and the eigenvectors $g_1, \dots, g_p$ of $\mathcal{S}$. The graphical representation of the PCs is obtained by plotting the first PC vs. the second (and eventually vs. the third).

↪ The components of the eigenvectors g_i are the weights of the original variables in the PCs.

10.3 Interpretation of the PCs

Recall that the main idea of PC transformations is to find the most informative projections that maximize variances. The most informative SLC is given by the first eigenvector. In Section 10.2 the eigenvectors were calculated for the bank data. In particular, with centered x's, we had:

Table 10.1 Proportion of variance of PC's

Eigenvalue	Proportion of variance	Cumulated proportion
2.985	0.67	0.67
0.931	0.21	0.88
0.242	0.05	0.93
0.194	0.04	0.97
0.085	0.02	0.99
0.035	0.01	1.00

$$y_1 = -0.044x_1 + 0.112x_2 + 0.139x_3 + 0.768x_4 + 0.202x_5 - 0.579x_6$$
$$y_2 = 0.011x_1 + 0.071x_2 + 0.066x_3 - 0.563x_4 + 0.659x_5 - 0.489x_6$$

and

$$\begin{aligned} x_1 &= \text{length}\\ x_2 &= \text{left height}\\ x_3 &= \text{right height}\\ x_4 &= \text{bottom frame}\\ x_5 &= \text{top frame}\\ x_6 &= \text{diagonal.} \end{aligned}$$

Hence, the first PC is essentially the difference between the bottom frame variable and the diagonal. The second PC is best described by the difference between the top frame variable and the sum of bottom frame and diagonal variables.

The weighting of the PCs tells us in which directions, expressed in original coordinates, the best variance explanation is obtained. A measure of how well the first q PCs explain variation is given by the relative proportion:

$$\psi_q = \frac{\sum_{j=1}^{q} \lambda_j}{\sum_{j=1}^{p} \lambda_j} = \frac{\sum_{j=1}^{q} \mathsf{Var}(Y_j)}{\sum_{j=1}^{p} \mathsf{Var}(Y_j)}. \tag{10.12}$$

Referring to the bank data Example 10.2, the (cumulative) proportions of explained variance are given in Table 10.1. The first PC ($q = 1$) already explains 67% of the variation. The first three ($q = 3$) PCs explain 93% of the variation. Once again it should be noted that PCs are not scale invariant, e.g., the PCs derived from the correlation matrix give different results than the PCs derived from the covariance matrix (see Section 10.5).

A good graphical representation of the ability of the PCs to explain the variation in the data is given by the scree plot shown in the lower right-hand window of Figure 10.3. The scree plot can be modified by using the relative proportions on the y-axis, as is shown in Figure 10.5 for the bank data set.

The covariance between the PC vector Y and the original vector X is calculated with the help of (10.4) as follows:

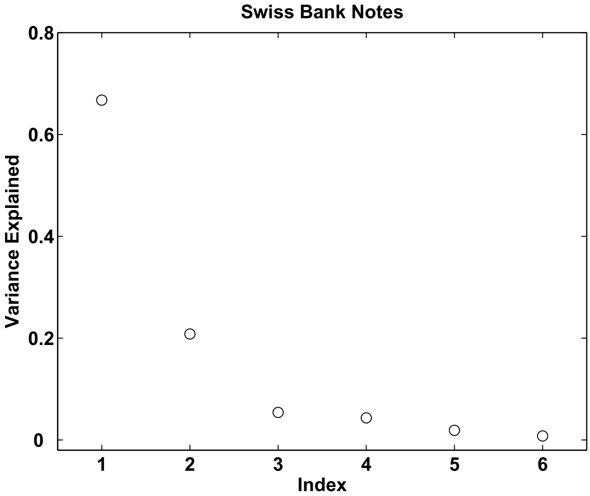


Fig. 10.5 Relative proportion of variance explained by PCs MVApcabanki

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathsf{E}(XY^\top) - \mathsf{E}\,X\,\mathsf{E}\,Y^\top = \mathsf{E}(XY^\top)\\ &= \mathsf{E}(XX^\top\Gamma) - \mu\mu^\top\Gamma = \mathrm{Var}(X)\Gamma\\ &= \Sigma\Gamma\\ &= \Gamma\Lambda\Gamma^\top\Gamma\\ &= \Gamma\Lambda. \end{aligned} \tag{10.13}$$

Hence, the correlation, $\rho_{X_iY_j}$, between variable X_i and the PC Y_j is

$$\rho_{X_iY_j} = \frac{\gamma_{ij}\lambda_j}{(\sigma_{X_iX_i}\lambda_j)^{1/2}} = \gamma_{ij}\left(\frac{\lambda_j}{\sigma_{X_iX_i}}\right)^{1/2}. \tag{10.14}$$

Using actual data, this of course translates into

$$r_{X_iY_j} = g_{ij}\left(\frac{\ell_j}{s_{X_iX_i}}\right)^{1/2}. \tag{10.15}$$

The correlations can be used to evaluate the relations between the PCs Y_j where $j = 1, \dots, q$, and the original variables X_i where $i = 1, \dots, p$. Note that

$$\sum_{j=1}^{p} r^2_{X_iY_j} = \frac{\sum_{j=1}^{p}\ell_j g_{ij}^2}{s_{X_iX_i}} = \frac{s_{X_iX_i}}{s_{X_iX_i}} = 1. \tag{10.16}$$

Indeed, $\sum_{j=1}^{p}\ell_j g_{ij}^2 = g_i^\top \mathcal{L} g_i$ is the (i,i)-element of the matrix $\mathcal{G}\mathcal{L}\mathcal{G}^\top = \mathcal{S}$, so that $r^2_{X_iY_j}$ may be seen as the proportion of variance of X_i explained by Y_j.

In the space of the first two PCs we plot these proportions, i.e., $r_{X_iY_1}$ versus $r_{X_iY_2}$. Figure 10.6 shows this for the bank notes example. This plot shows which of the original variables are most strongly correlated with PC Y_1 and Y_2.

From (10.16) it obviously follows that $r^2_{X_iY_1} + r^2_{X_iY_2} \le 1$ so that the points are always inside the circle of radius 1. In the bank notes example, the variables X_4, X_5

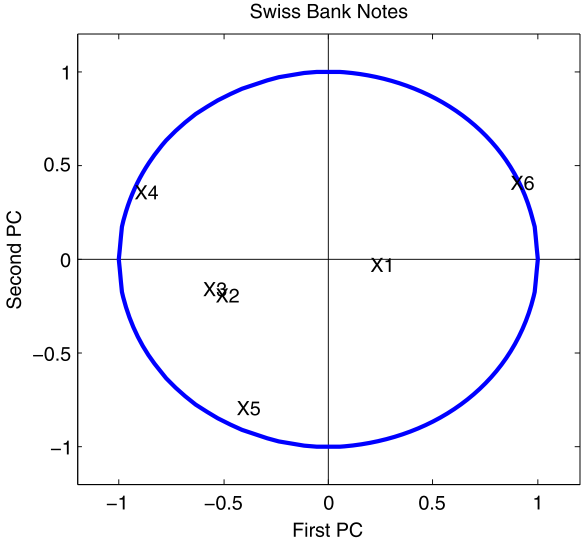


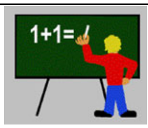
Fig. 10.6 The correlation of the original variable with the PCs MVApcabanki

Table 10.2 Correlation between the original variables and the PCs

	$r_{X_i Y_1}$	$r_{X_i Y_2}$	$r^2_{X_i Y_1} + r^2_{X_i Y_2}$
X_1 length	−0.201	0.028	0.041
X_2 left h.	0.538	0.191	0.326
X_3 right h.	0.597	0.159	0.381
X_4 lower	0.921	−0.377	0.991
X_5 upper	0.435	0.794	0.820
X_6 diagonal	−0.870	−0.410	0.926

and X_6 correspond to correlations near the periphery of the circle and are thus well explained by the first two PCs. Recall that we have interpreted the first PC as being essentially the difference between X_4 and X_6. This is also reflected in Figure 10.6 since the points corresponding to these variables lie on different sides of the vertical axis. An analogous remark applies to the second PC. We had seen that the second PC is well described by the difference between X_5 and the sum of X_4 and X_6. Now we are able to see this result again from Figure 10.6 since the point corresponding to X_5 lies above the horizontal axis and the points corresponding to X_4 and X_6 lie below.

The correlations of the original variables X_i and the first two PCs are given in Table 10.2 along with the cumulated percentage of variance of each variable explained by Y_1 and Y_2. This table confirms the above results. In particular, it confirms that the percentage of variance of X_1 (and X_2, X_3) explained by the first two PCs is relatively small and so are their weights in the graphical representation of the individual bank notes in the space of the first two PCs (as can be seen in the upper left plot in Figure 10.3). Looking simultaneously at Figure 10.6 and the upper left plot of Figure 10.3 shows that the genuine bank notes are roughly characterized by large values of X_6 and smaller values of X_4. The counterfeit bank notes show larger values of X_5 (see Example 7.15).



Summary

↪ The weighting of the PCs tells us in which directions, expressed in original coordinates, the best explanation of the variance is obtained. Note that the PCs are not scale invariant.

↪ A measure of how well the first q PCs explain variation is given by the relative proportion $\psi_q = \sum_{j=1}^{q} \lambda_j / \sum_{j=1}^{p} \lambda_j$. A good graphical representation of the ability of the PCs to explain the variation in the data is the scree plot of these proportions.

↪ The correlation between PC Y_j and an original variable X_i is $\rho_{X_i Y_j} = \gamma_{ij} \left(\frac{\lambda_j}{\sigma_{X_i X_i}}\right)^{1/2}$. For a data matrix this translates into $r^2_{X_i Y_j} = \frac{\ell_j g^2_{ij}}{s_{X_i X_i}}$. $r^2_{X_i Y_j}$ can be interpreted as the proportion of variance of X_i explained by Y_j. A plot of $r_{X_i Y_1}$ vs. $r_{X_i Y_2}$ shows which of the original variables are most strongly correlated with the PCs, namely those that are close to the periphery of the circle of radius 1.

10.4 Asymptotic Properties of the PCs

In practice, PCs are computed from sample data. The following theorem yields results on the asymptotic distribution of the sample PCs.

Theorem 10.4 *Let $\Sigma > 0$ with distinct eigenvalues, and let $\mathcal{U} \sim m^{-1} W_p(\Sigma, m)$ with spectral decompositions $\Sigma = \Gamma \Lambda \Gamma^\top$, and $\mathcal{U} = \mathcal{G} \mathcal{L} \mathcal{G}^\top$. Then*

(a) $\sqrt{m}(\ell - \lambda) \xrightarrow{\mathcal{L}} N_p(0, 2\Lambda^2)$, *where* $\ell = (\ell_1, \ldots, \ell_p)^\top$ *and* $\lambda = (\lambda_1, \ldots, \lambda_p)^\top$ *are the diagonals of* $\mathcal{L}$ *and* Λ,

(b) $\sqrt{m}(g_j - \gamma_j) \xrightarrow{\mathcal{L}} N_p(0, \mathcal{V}_j)$, *with* $\mathcal{V}_j = \lambda_j \sum_{k \neq j} \frac{\lambda_k}{(\lambda_k - \lambda_j)^2} \gamma_k \gamma_k^\top$,

(c) $\mathsf{Cov}(g_j, g_k) = \mathcal{V}_{jk}$, *where the* (r, s)*-element of the matrix* $\mathcal{V}_{jk}(p \times p)$ *is* $-\frac{\lambda_j \lambda_k \gamma_{rk} \gamma_{sj}}{m(\lambda_j - \lambda_k)^2}$,

(d) *the elements in* ℓ *are asymptotically independent of the elements in* $\mathcal{G}$.

Example 10.4 Since $n\mathcal{S} \sim W_p(\Sigma, n-1)$ if $X_1, \ldots, X_n$ are drawn from $N(\mu, \Sigma)$, we have that

$$\sqrt{n-1}(\ell_j - \lambda_j) \xrightarrow{\mathcal{L}} N(0, 2\lambda_j^2), \quad j = 1, \ldots, p. \tag{10.17}$$

Since the variance of (10.17) depends on the true mean λ_j a log transformation is useful. Consider $f(\ell_j) = \log(\ell_j)$. Then $\frac{d}{d\ell_j} f|_{\ell_j=\lambda_j} = \frac{1}{\lambda_j}$ and by the Transformation Theorem 4.11 we have from (10.17) that

$$\sqrt{n-1}(\log \ell_j - \log \lambda_j) \xrightarrow{\mathcal{L}} N(0, 2). \tag{10.18}$$

Hence,

$$\sqrt{\frac{n-1}{2}}(\log \ell_j - \log \lambda_j) \xrightarrow{\mathcal{L}} N(0, 1)$$

and a two-sided confidence interval at the $1 - \alpha = 0.95$ significance level is given by

$$\log(\ell_j) - 1.96\sqrt{\frac{2}{n-1}} \leq \log \lambda_j \leq \log(\ell_j) + 1.96\sqrt{\frac{2}{n-1}}.$$

In the bank data example we have that

$$\ell_1 = 2.98.$$

Therefore,

$$\log(2.98) \pm 1.96\sqrt{\frac{2}{199}} = \log(2.98) \pm 0.1965.$$

It can be concluded for the true eigenvalue that

$$\mathrm{P}\{\lambda_1 \in (2.448, 3.62)\} \approx 0.95.$$

Variance Explained by the First q PCs

The variance explained by the first q PCs is given by

$$\psi = \frac{\lambda_1 + \cdots + \lambda_q}{\sum_{j=1}^{p} \lambda_j}.$$

In practice this is estimated by

$$\widehat{\psi} = \frac{\ell_1 + \cdots + \ell_q}{\sum_{j=1}^{p} \ell_j}.$$

From Theorem 10.4 we know the distribution of $\sqrt{n-1}(\ell-\lambda)$. Since ψ is a nonlinear function of λ, we can again apply the Transformation Theorem 4.11 to obtain that

$$\sqrt{n-1}(\widehat{\psi}-\psi) \xrightarrow{\mathcal{L}} N(0, \mathcal{D}^\top \mathcal{V} \mathcal{D})$$

where $\mathcal{V} = 2\Lambda^2$ (from Theorem 10.4) and $\mathcal{D} = (d_1, \dots, d_p)^\top$ with

$$d_j = \frac{\partial \psi}{\partial \lambda_j} = \begin{cases} \frac{1-\psi}{\mathrm{tr}(\Sigma)} & \text{for } 1 \le j \le q, \\ \frac{-\psi}{\mathrm{tr}(\Sigma)} & \text{for } q+1 \le j \le p. \end{cases}$$

Given this result, the following theorem can be derived.

Theorem 10.5

$$\sqrt{n-1}(\widehat{\psi}-\psi) \xrightarrow{\mathcal{L}} N(0, \omega^2),$$

where

$$\begin{aligned} \omega^2 = \mathcal{D}^\top \mathcal{V} \mathcal{D} &= \frac{2}{\{\mathrm{tr}(\Sigma)\}^2} \left\{(1-\psi)^2(\lambda_1^2 + \cdots + \lambda_q^2) + \psi^2(\lambda_{q+1}^2 + \cdots + \lambda_p^2)\right\} \\ &= \frac{2\,\mathrm{tr}(\Sigma^2)}{\{\mathrm{tr}(\Sigma)\}^2}(\psi^2 - 2\beta\psi + \beta) \end{aligned}$$

and

$$\beta = \frac{\lambda_1^2 + \cdots + \lambda_q^2}{\lambda_1^2 + \cdots + \lambda_p^2}.$$

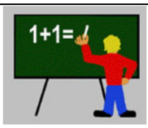
Example 10.5 From Section 10.3 it is known that the first PC for the Swiss bank notes resolves 67% of the variation. It can be tested whether the true proportion is actually 75%. Computing

$$\begin{aligned} \widehat{\beta} &= \frac{\ell_1^2}{\ell_1^2 + \cdots + \ell_p^2} = \frac{(2.985)^2}{(2.985)^2 + (0.931)^2 + \cdots (0.035)^2} = 0.902 \\ \mathrm{tr}(\mathcal{S}) &= 4.472 \\ \mathrm{tr}(\mathcal{S}^2) &= \sum_{j=1}^{p} \ell_j^2 = 9.883 \\ \widehat{\omega}^2 &= \frac{2\,\mathrm{tr}(\mathcal{S}^2)}{\{\mathrm{tr}(\mathcal{S})\}^2}(\widehat{\psi}^2 - 2\widehat{\beta}\widehat{\psi} + \widehat{\beta}) \\ &= \frac{2 \cdot 9.883}{(4.472)^2}\left\{(0.668)^2 - 2(0.902)(0.668) + 0.902\right\} = 0.142. \end{aligned}$$

Hence, a confidence interval at a significance of level $1-\alpha = 0.95$ is given by

$$0.668 \pm 1.96\sqrt{\frac{0.142}{199}} = (0.615, 0.720).$$

Clearly the hypothesis that $\psi = 75\%$ can be rejected!



Summary

$\hookrightarrow$ The eigenvalues ℓ_j and eigenvectors g_j are asymptotically, normally distributed, in particular $\sqrt{n-1}(\ell - \lambda) \xrightarrow{\mathcal{L}} N_p(0, 2\Lambda^2)$.

$\hookrightarrow$ For the eigenvalues it holds that $\sqrt{\frac{n-1}{2}}(\log \ell_j - \log \lambda_j) \xrightarrow{\mathcal{L}} N(0,1)$.

$\hookrightarrow$ Given an asymptotic, normal distribution approximate confidence intervals and tests can be constructed for the proportion of variance which is explained by the first q PCs. The two-sided confidence interval at the $1-\alpha = 0.95$ level is given by $\log(\ell_j) - 1.96\sqrt{\frac{2}{n-1}} \leq \log \lambda_j \leq \log(\ell_j) + 1.96\sqrt{\frac{2}{n-1}}$.

$\hookrightarrow$ It holds for $\widehat{\psi}$, the estimate of ψ (the proportion of the variance explained by the first q PCs) that $\sqrt{n-1}(\widehat{\psi} - \psi) \xrightarrow{\mathcal{L}} N(0, \omega^2)$, where ω is given in Theorem 10.5.

10.5 Normalized Principal Components Analysis

In certain situations the original variables can be heterogeneous w.r.t. their variances. This is particularly true when the variables are measured on heterogeneous scales (such as years, kilograms, dollars, ...). In this case a description of the information contained in the data needs to be provided which is robust w.r.t. the choice of scale. This can be achieved through a standardization of the variables, namely

$$\mathcal{X}_S = \mathcal{H}\mathcal{X}\mathcal{D}^{-1/2} \tag{10.19}$$

where $\mathcal{D} = \mathrm{diag}(s_{X_1X_1}, \ldots, s_{X_pX_p})$. Note that $\overline{x}_S = 0$ and $\mathcal{S}_{\mathcal{X}_S} = \mathcal{R}$, the correlation matrix of $\mathcal{X}$. The PC transformations of the matrix $\mathcal{X}_S$ are refereed to as the *Normalized Principal Components* (NPCs). The spectral decomposition of $\mathcal{R}$ is

$$\mathcal{R} = \mathcal{G}_{\mathcal{R}}\mathcal{L}_{\mathcal{R}}\mathcal{G}_{\mathcal{R}}^{\top}, \tag{10.20}$$

where $\mathcal{L}_{\mathcal{R}} = \mathrm{diag}(\ell_1^{\mathcal{R}}, \ldots, \ell_p^{\mathcal{R}})$ and $\ell_1^{\mathcal{R}} \geq \cdots \geq \ell_p^{\mathcal{R}}$ are the eigenvalues of $\mathcal{R}$ with corresponding eigenvectors $g_1^{\mathcal{R}}, \ldots, g_p^{\mathcal{R}}$ (note that here $\sum_{j=1}^{p} \ell_j^{\mathcal{R}} = \mathrm{tr}(\mathcal{R}) = p$).

The NPCs, Z_j, provide a representation of each individual, and is given by

$$\mathcal{Z} = \mathcal{X}_S \mathcal{G}_{\mathcal{R}} = (z_1, \ldots, z_p). \tag{10.21}$$

After transforming the variables, once again, we have that

$$\overline{z} = 0, \tag{10.22}$$

$$\mathcal{S}_{\mathcal{Z}} = \mathcal{G}_{\mathcal{R}}^{\top} \mathcal{S}_{\mathcal{X}_S} \mathcal{G}_{\mathcal{R}} = \mathcal{G}_{\mathcal{R}}^{\top} \mathcal{R} \mathcal{G}_{\mathcal{R}} = \mathcal{L}_{\mathcal{R}}. \tag{10.23}$$

The NPCs provide a perspective similar to that of the PCs, but in terms of the relative position of individuals, NPC gives each variable the same weight (with the PCs the variable with the largest variance received the largest weight).

Computing the covariance and correlation between X_i and Z_j is straightforward:

$$\mathcal{S}_{X_S,Z} = \frac{1}{n} \mathcal{X}_S^{\top} \mathcal{Z} = \mathcal{G}_{\mathcal{R}} \mathcal{L}_{\mathcal{R}}, \tag{10.24}$$

$$\mathcal{R}_{X_S,Z} = \mathcal{G}_{\mathcal{R}} \mathcal{L}_{\mathcal{R}} \mathcal{L}_{\mathcal{R}}^{-1/2} = \mathcal{G}_{\mathcal{R}} \mathcal{L}_{\mathcal{R}}^{1/2}. \tag{10.25}$$

The correlations between the original variables X_i and the NPCs Z_j are:

$$r_{X_i Z_j} = \sqrt{\ell_j} g_{R,ij} \tag{10.26}$$

$$\sum_{j=1}^{p} r_{X_i Z_j}^2 = 1 \tag{10.27}$$

(compare this to (10.15) and (10.16)). The resulting NPCs, the Z_j, can be interpreted in terms of the original variables and the role of each PC in explaining the variation in variable X_i can be evaluated.

10.6 Principal Components as a Factorial Method

The empirical PCs (normalized or not) turn out to be equivalent to the factors that one would obtain by decomposing the appropriate data matrix into its factors (see Chapter 9). It will be shown that the PCs are the factors representing the rows of the centered data matrix and that the NPCs correspond to the factors of the standardized data matrix. The representation of the columns of the standardized data matrix provides (at a scale factor) the correlations between the NPCs and the original variables. The derivation of the (N)PCs presented above will have a nice geometric justification here since they are the best fit in subspaces generated by the columns of the (transformed) data matrix $\mathcal{X}$. This analogy provides complementary interpretations of the graphical representations shown above.

Assume, as in Chapter 9, that we want to obtain representations of the individuals (the rows of $\mathcal{X}$) and of the variables (the columns of $\mathcal{X}$) in spaces of smaller dimension. To keep the representations simple, some prior transformations are performed.

Since the origin has no particular statistical meaning in the space of individuals, we will first shift the origin to the center of gravity, $\overline{x}$, of the point cloud. This is the same as analyzing the centered data matrix $\mathcal{X}_C = \mathcal{H}\mathcal{X}$. Now all of the variables have zero means, thus the technique used in Chapter 9 can be applied to the matrix $\mathcal{X}_C$. Note that the spectral decomposition of $\mathcal{X}_C^\top \mathcal{X}_C$ is related to that of $\mathcal{S}_X$, namely

$$\mathcal{X}_C^\top \mathcal{X}_C = \mathcal{X}^\top \mathcal{H}^\top \mathcal{H}\mathcal{X} = n\mathcal{S}_X = n\mathcal{G}\mathcal{L}\mathcal{G}^\top. \tag{10.28}$$

The factorial variables are obtained by projecting $\mathcal{X}_C$ on $\mathcal{G}$,

$$\mathcal{Y} = \mathcal{X}_C \mathcal{G} = (y_1, \ldots, y_p). \tag{10.29}$$

These are the same principal components obtained above, see formula (10.10). (Note that the y's here correspond to the z's in Section 9.2.) Since $\mathcal{H}\mathcal{X}_C = \mathcal{X}_C$, it immediately follows that

$$\overline{y} = 0, \tag{10.30}$$

$$\mathcal{S}_Y = \mathcal{G}^\top \mathcal{S}_X \mathcal{G} = \mathcal{L} = \operatorname{diag}(\ell_1, \ldots, \ell_p). \tag{10.31}$$

The scatterplot of the individuals on the factorial axes are thus centered around the origin and are more spread out in the first direction (first PC has variance ℓ_1) than in the second direction (second PC has variance ℓ_2).

The representation of the variables can be obtained using the Duality Relations (9.11), and (9.12). The projections of the columns of $\mathcal{X}_C$ onto the eigenvectors v_k of $\mathcal{X}_C \mathcal{X}_C^\top$ are

$$\mathcal{X}_C^\top v_k = \frac{1}{\sqrt{n\ell_k}} \mathcal{X}_C^\top \mathcal{X}_C g_k = \sqrt{n\ell_k}\, g_k. \tag{10.32}$$

Thus the projections of the variables on the first p axes are the columns of the matrix

$$\mathcal{X}_C^\top \mathcal{V} = \sqrt{n}\,\mathcal{G}\mathcal{L}^{1/2}. \tag{10.33}$$

Considering the geometric representation, there is a nice statistical interpretation of the angle between two columns of $\mathcal{X}_C$. Given that

$$x_{C[j]}^\top x_{C[k]} = n s_{X_j X_k}, \tag{10.34}$$

$$||x_{C[j]}||^2 = n s_{X_j X_j}, \tag{10.35}$$

where $x_{C[j]}$ and $x_{C[k]}$ denote the j-th and k-th column of $\mathcal{X}_C$, it holds that in the full space of the variables, if θ_{jk} is the angle between two variables, $x_{C[j]}$ and $x_{C[k]}$, then

$$\cos\theta_{jk} = \frac{x_{C[j]}^\top x_{C[k]}}{\|x_{C[j]}\|\;\|x_{C[k]}\|} = r_{X_j X_k} \tag{10.36}$$

(Example 2.11 shows the general connection that exists between the angle and correlation of two variables). As a result, the relative positions of the variables in the scatterplot of the first columns of $\mathcal{X}_C^\top \mathcal{V}$ may be interpreted in terms of their correlations; the plot provides a picture of the correlation structure of the original data set.

Clearly, one should take into account the percentage of variance explained by the chosen axes when evaluating the correlation.

The NPCs can also be viewed as a factorial method for reducing the dimension. The variables are again standardized so that each one has mean zero and unit variance and is independent of the scale of the variables. The factorial analysis of $\mathcal{X}_S$ provides the NPCs. The spectral decomposition of $\mathcal{X}_S^\top \mathcal{X}_S$ is related to that of $\mathcal{R}$, namely

$$\mathcal{X}_S^\top \mathcal{X}_S = \mathcal{D}^{-1/2}\mathcal{X}^\top \mathcal{H}\mathcal{X}\mathcal{D}^{-1/2} = n\mathcal{R} = n\mathcal{G}_\mathcal{R}\mathcal{L}_\mathcal{R}\mathcal{G}_\mathcal{R}^\top.$$

The NPCs Z_j, given by (10.21), may be viewed as the projections of the rows of $\mathcal{X}_S$ onto $\mathcal{G}_R$.

The representation of the variables are again given by the columns of

$$\mathcal{X}_S^\top \mathcal{V}_\mathcal{R} = \sqrt{n}\mathcal{G}_\mathcal{R}\mathcal{L}_\mathcal{R}^{1/2}. \tag{10.37}$$

Comparing (10.37) and (10.25) we see that the projections of the variables in the factorial analysis provide the correlation between the NPCs $\mathcal{Z}_k$ and the original variables $x_{[j]}$ (up to the factor $\sqrt{n}$ which could be the scale of the axes).

This implies that a deeper interpretation of the representation of the individuals can be obtained by looking simultaneously at the graphs plotting the variables. Note that

$$x_{S[j]}^\top x_{S[k]} = n r_{X_j X_k}, \tag{10.38}$$

$$\|x_{S[j]}\|^2 = n, \tag{10.39}$$

where $x_{S[j]}$ and $x_{S[k]}$ denote the j-th and k-th column of $\mathcal{X}_S$. Hence, in the full space, all the standardized variables (columns of $\mathcal{X}_S$) are contained within the "sphere" in $\mathbb{R}^n$, which is centered at the origin and has radius $\sqrt{n}$ (the scale of the graph). As in (10.36), given the angle θ_{jk} between two columns $x_{S[j]}$ and $x_{S[k]}$, it holds that

$$\cos\theta_{jk} = r_{X_j X_k}. \tag{10.40}$$

Therefore, when looking at the representation of the variables in the spaces of reduced dimension (for instance the first two factors), we have a picture of the correlation structure between the original X_i's in terms of their angles. Of course, the quality of the representation in those subspaces has to be taken into account, which is presented in the next section.

Quality of the Representations

As said before, an overall measure of the quality of the representation is given by

$$\psi = \frac{\ell_1 + \ell_2 + \cdots + \ell_q}{\sum_{j=1}^p \ell_j}.$$

In practice, q is chosen to be equal to 1, 2 or 3. Suppose for instance that $\psi = 0.93$ for $q = 2$. This means that the graphical representation in two dimensions captures 93% of the total variance. In other words, there is minimal dispersion in a third direction (no more than 7%).

It can be useful to check if each individual is well represented by the PCs. Clearly, the proximity of two individuals on the projected space may not necessarily coincide with the proximity in the full original space $\mathbb{R}^p$, which may lead to erroneous interpretations of the graphs. In this respect, it is worth computing the angle ϑ_{ik} between the representation of an individual i and the k-th PC or NPC axis. This can be done using (2.40), i.e.,

$$\cos \vartheta_{ik} = \frac{y_i^\top e_k}{\|y_i\| \|e_k\|} = \frac{y_{ik}}{\|x_{Ci}\|}$$

for the PCs or analogously

$$\cos \zeta_{ik} = \frac{z_i^\top e_k}{\|z_i\| \|e_k\|} = \frac{z_{ik}}{\|x_{Si}\|}$$

for the NPCs, where e_k denotes the k-th unit vector $e_k = (0, \ldots, 1, \ldots, 0)^\top$. An individual i will be represented on the k-th PC axis if its corresponding angle is small, i.e., if $\cos^2 \vartheta_{ik}$ for $k = 1, \ldots, p$ is close to one. Note that for each individual i,

$$\sum_{k=1}^{p} \cos^2 \vartheta_{ik} = \frac{y_i^\top y_i}{x_{Ci}^\top x_{Ci}} = \frac{x_{Ci}^\top \mathcal{G}\mathcal{G}^\top x_{Ci}}{x_{Ci}^\top x_{Ci}} = 1.$$

The values $\cos^2 \vartheta_{ik}$ are sometimes called the relative contributions of the k-th axis to the representation of the i-th individual, e.g., if $\cos^2 \vartheta_{i1} + \cos^2 \vartheta_{i2}$ is large (near one), we know that the individual i is well represented on the plane of the first two principal axes since its corresponding angle with the plane is close to zero.

We already know that the quality of the representation of the variables can be evaluated by the percentage of X_i's variance that is explained by a PC, which is given by $r^2_{X_i Y_j}$ or $r^2_{X_i Z_j}$ according to (10.16) and (10.27) respectively.

Example 10.6 Let us return to the French food expenditure example, see Appendix B.6. This yields a two-dimensional representation of the individuals as shown in Figure 10.7.

Calculating the matrix $\mathcal{G}_\mathcal{R}$ we have

$$\mathcal{G}_\mathcal{R} = \begin{pmatrix} -0.240 & 0.622 & -0.011 & -0.544 & 0.036 & 0.508 \\ -0.466 & 0.098 & -0.062 & -0.023 & -0.809 & -0.301 \\ -0.446 & -0.205 & 0.145 & 0.548 & -0.067 & 0.625 \\ -0.462 & -0.141 & 0.207 & -0.053 & 0.411 & -0.093 \\ -0.438 & -0.197 & 0.356 & -0.324 & 0.224 & -0.350 \\ -0.281 & 0.523 & -0.444 & 0.450 & 0.341 & -0.332 \\ 0.206 & 0.479 & 0.780 & 0.306 & -0.069 & -0.138 \end{pmatrix},$$

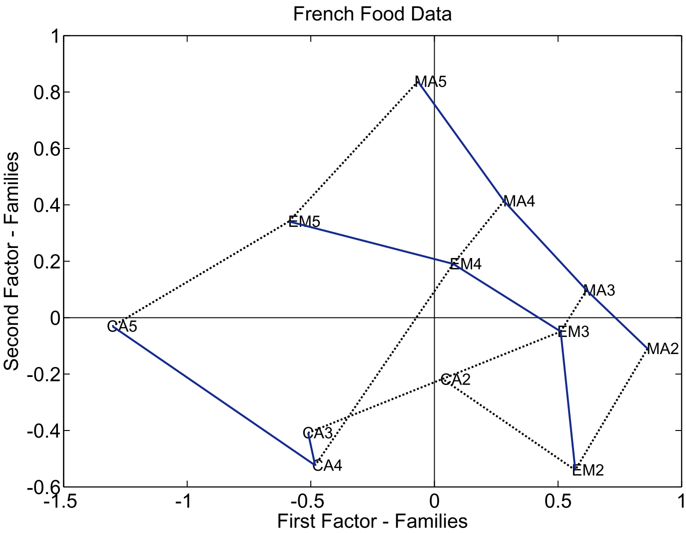


Fig. 10.7 Representation of the individuals MVAnpcafood

Table 10.3 Eigenvalues and explained variance

eigenvalues	proportion of variance	cumulated proportion
4.333	0.6190	61.9
1.830	0.2620	88.1
0.631	0.0900	97.1
0.128	0.0180	98.9
0.058	0.0080	99.7
0.019	0.0030	99.9
0.001	0.0001	100.0

which gives the weights of the variables (milk, vegetables, etc.). The eigenvalues ℓ_j and the proportions of explained variance are given in Table 10.3.

The interpretation of the principal components are best understood when looking at the correlations between the original X_i's and the PCs. Since the first two PCs explain 88.1% of the variance, we limit ourselves to the first two PCs. The results are shown in Table 10.4. The two-dimensional graphical representation of the variables in Figure 10.8 is based on the first two columns of Table 10.4.

The plots are the projections of the variables into $\mathbb{R}^2$. Since the quality of the representation is good for all the variables (except maybe X_7), their relative angles give a picture of their original correlation: wine is negatively correlated with the vegetables, fruits, meat and poultry groups ($\theta > 90°$), whereas taken individually this latter grouping of variables are highly positively correlated with each other ($\theta \approx 0$). Bread and milk are positively correlated but poorly correlated with meat, fruits and poultry ($\theta \approx 90°$).

Table 10.4 Correlations with PCs

	$r_{X_i Z_1}$	$r_{X_i Z_2}$	$r^2_{X_i Z_1} + r^2_{X_i Z_2}$
X_1: bread	−0.499	0.842	0.957
X_2: vegetables	−0.970	0.133	0.958
X_3: fruits	−0.929	−0.278	0.941
X_4: meat	−0.962	−0.191	0.962
X_5: poultry	−0.911	−0.266	0.901
X_6: milk	−0.584	0.707	0.841
X_7: wine	0.428	0.648	0.604

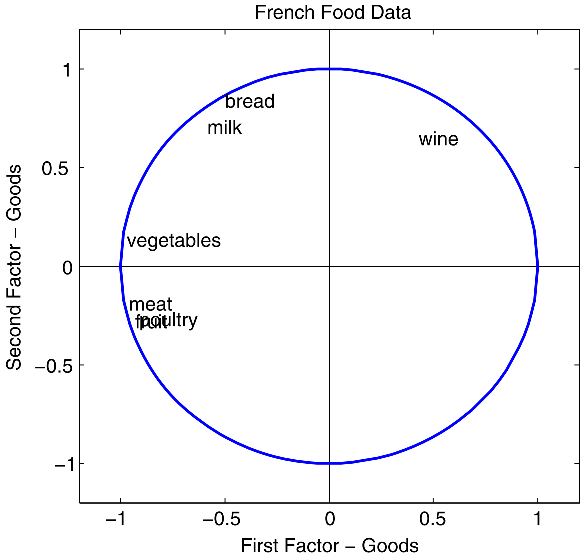
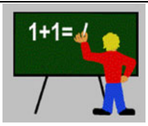


Fig. 10.8 Representation of the variables MVAnpcafood

Now the representation of the individuals in Figure 10.7 can be interpreted better. From Figure 10.8 and Table 10.4 we can see that the the first factor Z_1 is a vegetable–meat–poultry–fruit factor (with a negative sign), whereas the second factor is a milk–bread–wine factor (with a positive sign). Note that this corresponds to the most important weights in the first columns of $\mathcal{G}_{\mathcal{R}}$. In Figure 10.7 lines were drawn to connect families of the same size and families of the same professional types. A grid can clearly be seen (with a slight deformation by the manager families) that shows the families with higher expenditures (higher number of children) on the left.

Considering both figures together explains what types of expenditures are responsible for similarities in food expenditures. Bread, milk and wine expenditures are similar for manual workers and employees. Families of managers are characterized by higher expenditures on vegetables, fruits, meat and poultry. Very often when analyzing NPCs (and PCs), it is illuminating to use such a device to introduce qualitative aspects of individuals in order to enrich the interpretations of the graphs.



Summary

↪ NPCs are PCs applied to the standardized (normalized) data matrix $\mathcal{X}_S$.

↪ The graphical representation of NPCs provides a similar type of picture as that of PCs, the difference being in the relative position of individuals, i.e., each variable in NPCs has the same weight (in PCs, the variable with the largest variance has the largest weight).

↪ The quality of the representation is evaluated by $\psi = (\sum_{j=1}^{p} \ell_j)^{-1}(\ell_1 + \ell_2 + \cdots + \ell_q)$.

↪ The quality of the representation of a variable can be evaluated by the percentage of X_i's variance that is explained by a PC, i.e., $r^2_{X_i Y_j}$.

10.7 Common Principal Components

In many applications a statistical analysis is simultaneously done for groups of data. In this section a technique is presented that allows us to analyze group elements that have common PCs. From a statistical point of view, estimating PCs simultaneously in different groups will result in a joint dimension reducing transformation. This multi-group PCA, the so called common principle components analysis (CPCA), yields the joint eigenstructure across groups.

In addition to traditional PCA, the basic assumption of CPCA is that the space spanned by the eigenvectors is identical *across* several groups, whereas variances associated with the components are allowed to vary.

More formally, the hypothesis of common principle components can be stated in the following way (Flury, 1988):

$$H_{CPC} : \Sigma_i = \Gamma \Lambda_i \Gamma^\top, \quad i = 1, \ldots, k$$

where Σ_i is a positive definite $p \times p$ population covariance matrix for every i, $\Gamma = (\gamma_1, \ldots, \gamma_p)$ is an orthogonal $p \times p$ transformation matrix and $\Lambda_i = \mathrm{diag}(\lambda_{i1}, \ldots, \lambda_{ip})$ is the matrix of eigenvalues. Moreover, assume that all λ_i are distinct.

Let $\mathcal{S}$ be the (unbiased) sample covariance matrix of an underlying p-variate normal distribution $N_p(\mu, \Sigma)$ with sample size n. Then the distribution of $n\mathcal{S}$ has $n-1$ degrees of freedom and is known as the Wishart distribution (Muirhead, 1982, p. 86):

$$n\mathcal{S} \sim \mathcal{W}_p(\Sigma, n-1).$$

The density is given in (5.16). Hence, for a given Wishart matrix $\mathcal{S}_i$ with sample size n_i, the likelihood function can be written as

$$L(\Sigma_1, \dots, \Sigma_k) = C \prod_{i=1}^{k} \exp\left[\operatorname{tr}\left\{-\frac{1}{2}(n_i - 1)\Sigma_i^{-1}\mathcal{S}_i\right\}\right] |\Sigma_i|^{-\frac{1}{2}(n_i-1)} \qquad (10.41)$$

where C is a constant independent of the parameters Σ_i. Maximizing the likelihood is equivalent to minimizing the function

$$g(\Sigma_1, \dots, \Sigma_k) = \sum_{i=1}^{k}(n_i - 1)\left\{\log|\Sigma_i| + \operatorname{tr}(\Sigma_i^{-1}\mathcal{S}_i)\right\}.$$

Assuming that H_{CPC} holds, i.e., in replacing Σ_i by $\Gamma\Lambda_i\Gamma^\top$, after some manipulations one obtains

$$g(\Gamma, \Lambda_1, \dots, \Lambda_k) = \sum_{i=1}^{k}(n_i - 1)\sum_{j=1}^{p}\left(\log\lambda_{ij} + \frac{\gamma_j^\top \mathcal{S}_i \gamma_j}{\lambda_{ij}}\right).$$

As we know from Section 2.2, the vectors γ_j in Γ have to be orthogonal. Orthogonality of the vectors γ_j is achieved using the Lagrange method, i.e., we impose the p constraints $\gamma_j^\top \gamma_j = 1$ using the Lagrange multipliers μ_j, and the remaining $p(p-1)/2$ constraints $\gamma_h^\top \gamma_j = 0$ for $h \neq j$ using the multiplier $2\mu_{hj}$ (Flury, 1988). This yields

$$g^*(\Gamma, \Lambda_1, \dots, \Lambda_k) = g(\cdot) - \sum_{j=1}^{p}\mu_j(\gamma_j^\top\gamma_j - 1) - 2\sum_{h=1}^{p}\sum_{j=h+1}^{p}\mu_{hj}\gamma_h^\top\gamma_j.$$

Taking partial derivatives with respect to all λ_{im} and γ_m, it can be shown that the solution of the CPC model is given by the generalized system of characteristic equations

$$\gamma_m^\top\left\{\sum_{i=1}^{k}(n_i - 1)\frac{\lambda_{im} - \lambda_{ij}}{\lambda_{im}\lambda_{ij}}\mathcal{S}_i\right\}\gamma_j = 0, \quad m, j = 1, \dots, p,\ m \neq j. \qquad (10.42)$$

This system can be solved using

$$\lambda_{im} = \gamma_m^\top \mathcal{S}\gamma_m, \quad i = 1, \dots, k,\ m = 1, \dots, p$$

under the constraints

$$\gamma_m^\top\gamma_j = \begin{cases} 0 & m \neq j \\ 1 & m = j. \end{cases}$$

Flury (1988) proves existence and uniqueness of the maximum of the likelihood function, and Flury and Gautschi (1986) provide a numerical algorithm.

Example 10.7 As an example we provide the data sets XFGvolsurf01, XFGvolsurf02 and XFGvolsurf03 that have been used in Fengler, Härdle and Villa (2003) to estimate common principle components for the implied volatility surfaces of

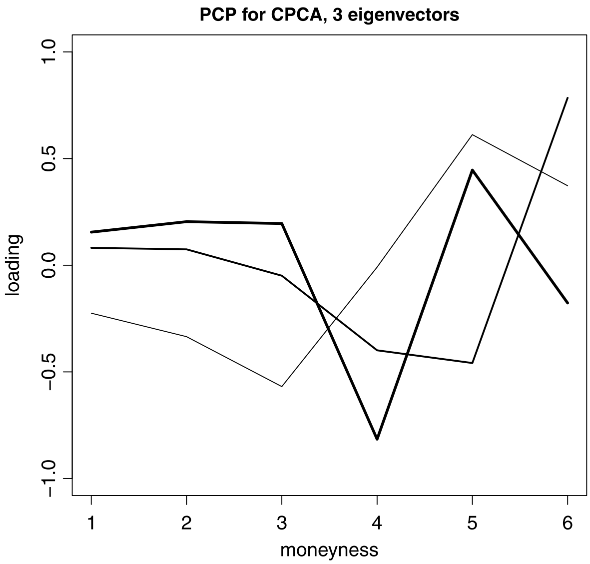


Fig. 10.9 Factor loadings of the first (thick), the second (medium), and the third (thin) PC `MVAcpcaiv`

the DAX 1999. The data has been generated by smoothing an implied volatility surface day by day. Next, the estimated grid points have been grouped into maturities of $\tau = 1$, $\tau = 2$ and $\tau = 3$ months and transformed into a vector of time series of the "smile", i.e., each element of the vector belongs to a distinct moneyness ranging from 0.85 to 1.10.

Figure 10.9 shows the first three eigenvectors in a parallel coordinate plot. The basic structure of the first three eigenvectors is not altered. We find a shift, a slope and a twist structure. This structure is *common* to all maturity groups, i.e., when exploiting PCA as a dimension reducing tool, the same transformation applies to each group! However, by comparing the size of eigenvalues among groups we find that variability is decreasing across groups as we move from the short term contracts to long term contracts.

Before drawing conclusions we should convince ourselves that the CPC model is truly a good description of the data. This can be done by using a likelihood ratio test. The likelihood ratio statistic for comparing a restricted (the CPC) model against the unrestricted model (the model where all covariances are treated separately) is given by

$$T_{(n_1,n_2,\ldots,n_k)} = -2\log\frac{L(\widehat{\Sigma}_1,\ldots,\widehat{\Sigma}_k)}{L(\mathcal{S}_1,\ldots,\mathcal{S}_k)}.$$

Inserting the likelihood function, we find that this is equivalent to

$$T_{(n_1,n_2,\ldots,n_k)} = \sum_{i=1}^{k}(n_i-1)\frac{\det\,(\widehat{\Sigma}_i)}{\det\,(\mathcal{S}_i)},$$

which has a χ^2 distribution as $\min(n_i)$ tends to infinity with

$$k\left\{\frac{1}{2}p(p-1)+1\right\}-\left\{\frac{1}{2}p(p-1)+kp\right\}=\frac{1}{2}(k-1)p(p-1)$$

degrees of freedom. This test is included in the quantlet `MVAcpcaiv`.

The calculations yield $T_{(n_1,n_2,\ldots,n_k)} = 31.836$, which corresponds to the p-value $p = 0.37512$ for the $\chi^2(30)$ distribution. Hence we cannot reject the CPC model against the unrestricted model, where PCA is applied to each maturity separately.

Using the methods in Section 10.3, we can estimate the amount of variability, ζ_l, explained by the first l principal components: (only a few factors, three at the most, are needed to capture a large amount of the total variability present in the data). Since the model now captures the variability in both the strike and maturity dimensions, this is a suitable starting point for a simplified VaR calculation for delta-gamma neutral option portfolios using Monte Carlo methods, and is hence a valuable insight in risk management.

10.8 Boston Housing

A set of transformations were defined in Chapter 1 for the Boston Housing data set that resulted in "regular" marginal distributions. The usefulness of principal component analysis with respect to such high-dimensional data sets will now be shown. The variable X_4 is dropped because it is a discrete 0–1 variable. It will be used later, however, in the graphical representations. The scale difference of the remaining 13 variables motivates a NPCA based on the correlation matrix.

The eigenvalues and the percentage of explained variance are given in Table 10.5.

Table 10.5 Eigenvalues and percentage of explained variance for Boston housing data `MVAnpcahousi`

Eigenvalue	Percentages	Cumulated percentages
7.2852	0.5604	0.5604
1.3517	0.1040	0.6644
1.1266	0.0867	0.7510
0.7802	0.0600	0.8111
0.6359	0.0489	0.8600
0.5290	0.0407	0.9007
0.3397	0.0261	0.9268
0.2628	0.0202	0.9470
0.1936	0.0149	0.9619
0.1547	0.0119	0.9738
0.1405	0.0108	0.9846
0.1100	0.0085	0.9931
0.0900	0.0069	1.0000

Table 10.6 Correlations of the first three PC's with the original variables
MVAnpcahous

	PC_1	PC_2	PC_3
X_1	−0.9076	0.2247	0.1457
X_2	0.6399	−0.0292	0.5058
X_3	−0.8580	0.0409	−0.1845
X_5	−0.8737	0.2391	−0.1780
X_6	0.5104	0.7037	0.0869
X_7	−0.7999	0.1556	−0.2949
X_8	0.8259	−0.2904	0.2982
X_9	−0.7531	0.2857	0.3804
X_{10}	−0.8114	0.1645	0.3672
X_{11}	−0.5674	−0.2667	0.1498
X_{12}	0.4906	−0.1041	−0.5170
X_{13}	−0.7996	−0.4253	−0.0251
X_{14}	0.7366	0.5160	−0.1747

The first principal component explains 56% of the total variance and the first three components together explain more than 75%. These results imply that it is sufficient to look at 2, maximum 3, principal components.

Table 10.6 provides the correlations between the first three PC's and the original variables. These can be seen in Figure 10.10.

The correlations with the first PC show a very clear pattern. The variables X_2, X_6, X_8, X_{12}, and X_{14} are strongly positively correlated with the first PC, whereas the remaining variables are highly negatively correlated. The minimal correlation in the absolute value is 0.5. The first PC axis could be interpreted as a quality of life and house indicator. The second axis, given the polarities of X_{11} and X_{13} and of X_6 and X_{14}, can be interpreted as a social factor explaining only 10% of the total variance. The third axis is dominated by a polarity between X_2 and X_{12}.

The set of individuals from the first two PCs can be graphically interpreted if the plots are color coded with respect to some particular variable of interest. Figure 10.11 color codes $X_{14} >$ median as red points. Clearly the first and second PCs are related to house value. The situation is less clear in Figure 10.12 where the color code corresponds to X_4, the Charles River indicator, i.e., houses near the river are colored red.

10.9 More Examples

Example 10.8 Let us now apply the PCA to the *standardized* bank data set (Table B.2). Figure 10.13 shows some PC plots of the bank data set. The genuine and counterfeit bank notes are marked by "o" and "+" respectively.

The vector of eigenvalues of $\mathcal{R}$ is

$$\ell = (2.946, 1.278, 0.869, 0.450, 0.269, 0.189)^\top .$$

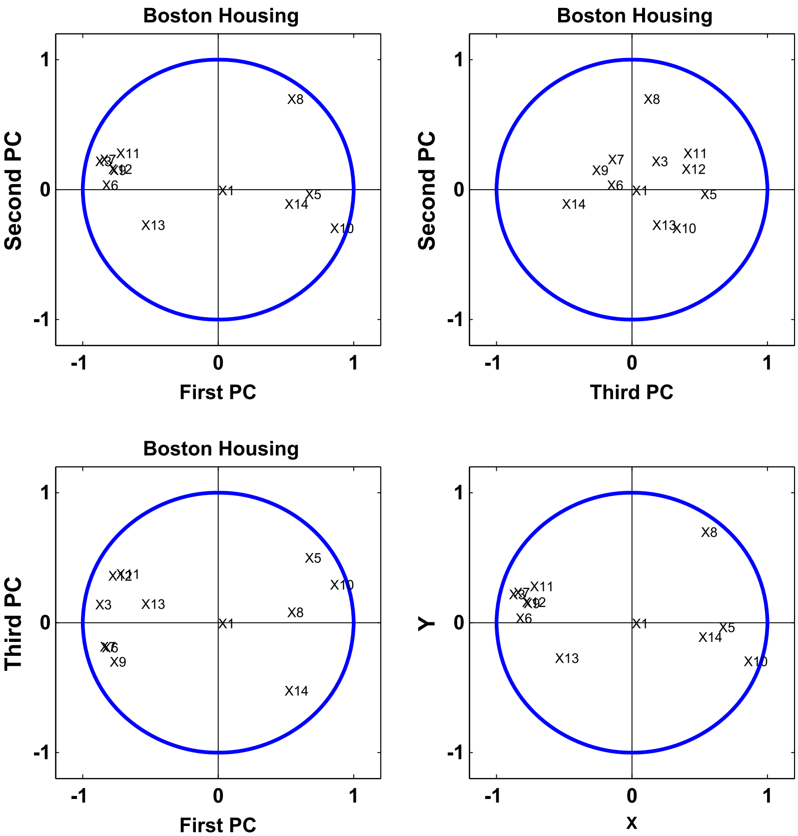


Fig. 10.10 NPCA for the Boston housing data, correlations of first three PCs with the original variables MVAnpcahousi

The eigenvectors g_j are given by the columns of the matrix

$$\mathcal{G} = \begin{pmatrix} -0.007 & -0.815 & 0.018 & 0.575 & 0.059 & 0.031 \\ 0.468 & -0.342 & -0.103 & -0.395 & -0.639 & -0.298 \\ 0.487 & -0.252 & -0.123 & -0.430 & 0.614 & 0.349 \\ 0.407 & 0.266 & -0.584 & 0.404 & 0.215 & -0.462 \\ 0.368 & 0.091 & 0.788 & 0.110 & 0.220 & -0.419 \\ -0.493 & -0.274 & -0.114 & -0.392 & 0.340 & -0.632 \end{pmatrix}.$$

Each original variable has the same weight in the analysis and the results are independent of the scale of each variable.

The proportions of explained variance are given in Table 10.7. It can be concluded that the representation in two dimensions should be sufficient. The correlations leading to Figure 10.14 are given in Table 10.8. The picture is different from the one obtained in Section 10.3 (see Table 10.2). Here, the first factor is mainly

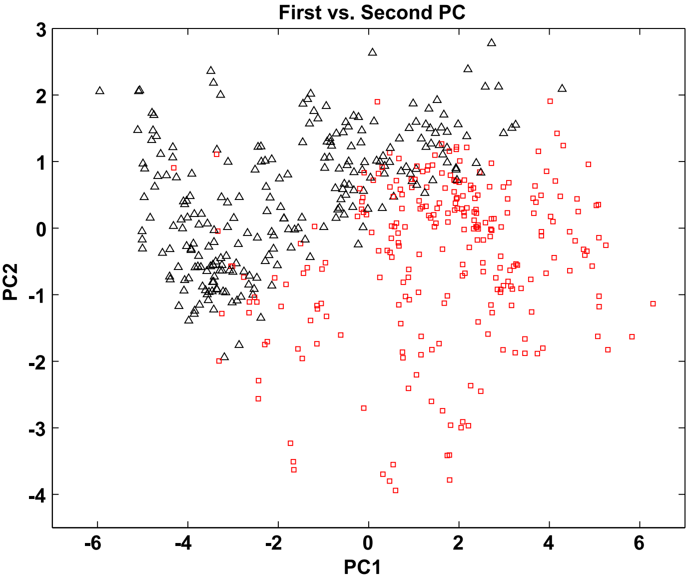


Fig. 10.11 NPC analysis for the Boston housing data, scatterplot of the first two PCs. More expensive houses are marked with red color MVAnpcahous

Table 10.7 Eigenvalues and proportions of explained variance

ℓ_j	Proportion of variances	Cumulated proportion
2.946	0.491	49.1
1.278	0.213	70.4
0.869	0.145	84.9
0.450	0.075	92.4
0.264	0.045	96.9
0.189	0.032	100.0

Table 10.8 Correlations with PCs

	$r_{X_i Z_1}$	$r_{X_i Z_2}$	$r^2_{X_i Z_1} + r^2_{X_i Z_2}$
X_1: length	-0.012	-0.922	0.85
X_2: left height	0.803	-0.387	0.79
X_3: right height	0.835	-0.285	0.78
X_4: lower	0.698	0.301	0.58
X_5: upper	0.631	0.104	0.41
X_6: diagonal	-0.847	-0.310	0.81

a left–right vs. diagonal factor and the second one is a length factor (with negative weight). Take another look at Figure 10.13, where the individual bank notes are displayed. In the upper left graph it can be seen that the genuine bank notes

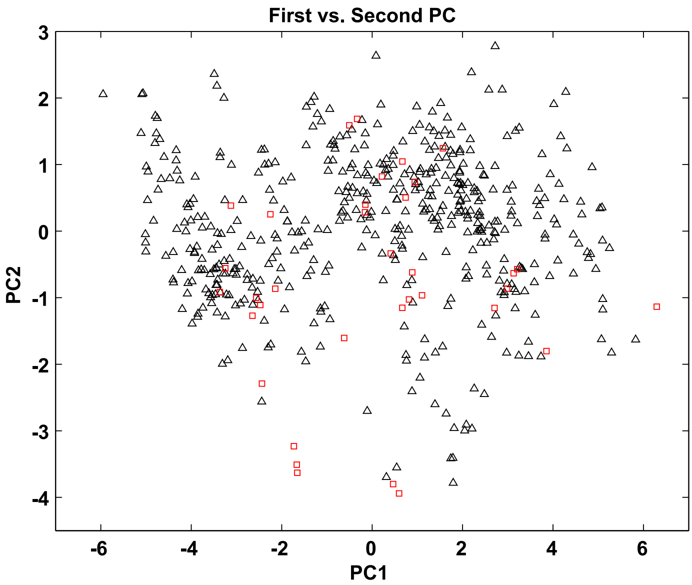


Fig. 10.12 NPC analysis for the Boston housing data, scatterplot of the first two PCs. Houses close to the Charles River are indicated with red squares MVAnpcahous

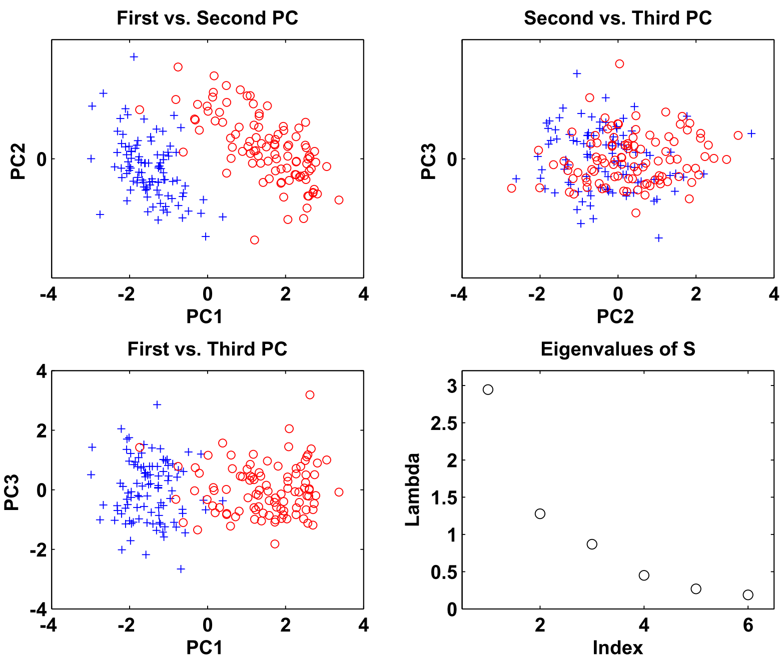


Fig. 10.13 Principal components of the *standardized* bank data MVAnpcabank

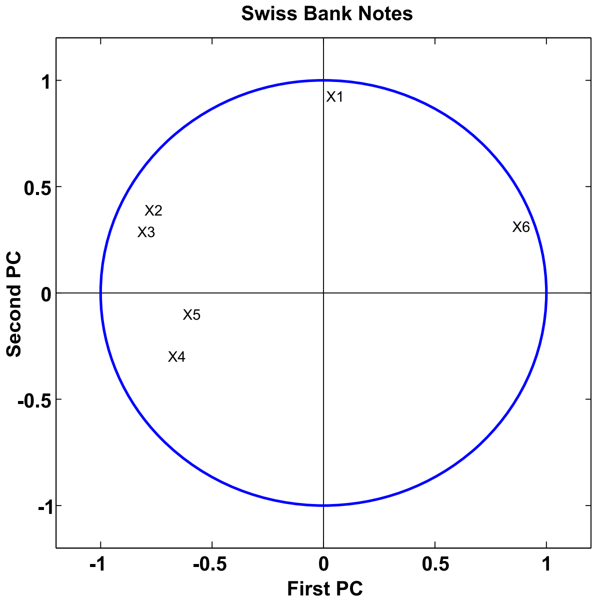


Fig. 10.14 The correlations of the original variable with the PCs MVAnpcabanki

are for the most part in the south-eastern portion of the graph featuring a larger diagonal, smaller height ($Z_1 < 0$) and also a larger length ($Z_2 < 0$). Note also that Figure 10.14 gives an idea of the correlation structure of the original data matrix.

Example 10.9 Consider the data of 79 U.S. companies given in Table B.5. The data is first standardized by subtracting the mean and dividing by the standard deviation. Note that the data set contains six variables: assets (X_1), sales (X_2), market value (X_3), profits (X_4), cash flow (X_5), number of employees (X_6).

Calculating the corresponding vector of eigenvalues gives

$$\ell = (5.039, 0.517, 0.359, 0.050, 0.029, 0.007)^\top$$

and the matrix of eigenvectors is

$$\mathcal{G} = \begin{pmatrix} 0.340 & -0.849 & -0.339 & 0.205 & 0.077 & -0.006 \\ 0.423 & -0.170 & 0.379 & -0.783 & -0.006 & -0.186 \\ 0.434 & 0.190 & -0.192 & 0.071 & -0.844 & 0.149 \\ 0.420 & 0.364 & -0.324 & 0.156 & 0.261 & -0.703 \\ 0.428 & 0.285 & -0.267 & -0.121 & 0.452 & 0.667 \\ 0.397 & 0.010 & 0.726 & 0.548 & 0.098 & 0.065 \end{pmatrix}.$$

Using this information the graphical representations of the first two principal components are given in Figure 10.15. The different sectors are marked by the following symbols:

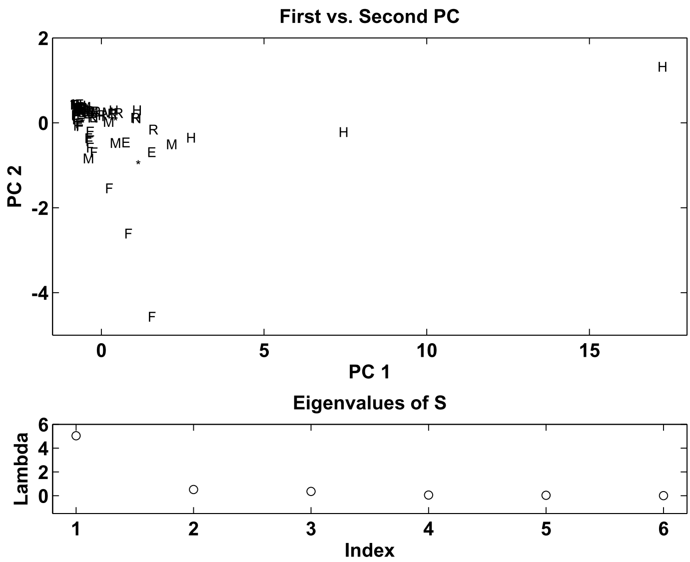


Fig. 10.15 Principal components of the U.S. company data MVAnpcausco

H	...	Hi Tech and Communication
E	...	Energy
F	...	Finance
M	...	Manufacturing
R	...	Retail
$\star$	...	all other sectors.

The two outliers in the right-hand side of the graph are IBM and General Electric (GE), which differ from the other companies with their high market values. As can be seen in the first column of $\mathcal{G}$, market value has the largest weight in the first PC, adding to the isolation of these two companies. If IBM and GE were to be excluded from the data set, a completely different picture would emerge, as shown in Figure 10.16. In this case the vector of eigenvalues becomes

$$\ell = (3.191, 1.535, 0.791, 0.292, 0.149, 0.041)^\top,$$

and the corresponding matrix of eigenvectors is

$$\mathcal{G} = \begin{pmatrix} 0.263 & -0.408 & -0.800 & -0.067 & 0.333 & 0.099 \\ 0.438 & -0.407 & 0.162 & -0.509 & -0.441 & -0.403 \\ 0.500 & -0.003 & -0.035 & 0.801 & -0.264 & -0.190 \\ 0.331 & 0.623 & -0.080 & -0.192 & 0.426 & -0.526 \\ 0.443 & 0.450 & -0.123 & -0.238 & -0.335 & 0.646 \\ 0.427 & -0.277 & 0.558 & 0.021 & 0.575 & 0.313 \end{pmatrix}.$$

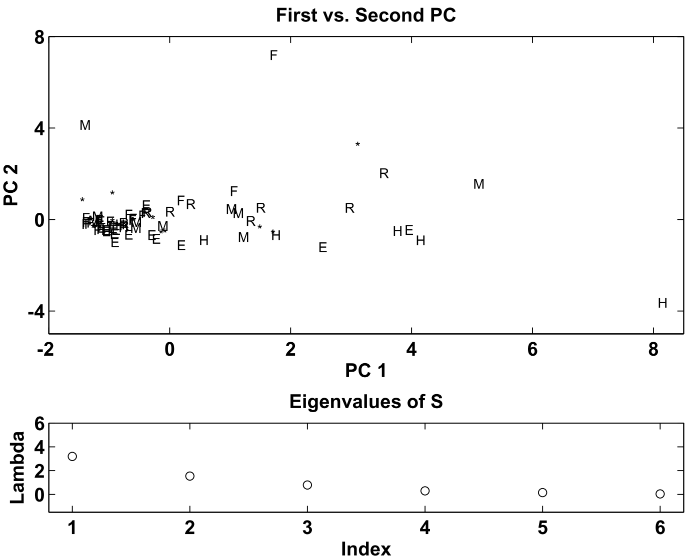


Fig. 10.16 Principal components of the U.S. company data (without IBM and General Electric) MVAnpcausco2

Table 10.9 Eigenvalues and proportions of explained variance

ℓ_j	Proportion of variance	Cumulated proportion
3.191	0.532	0.532
1.535	0.256	0.788
0.791	0.132	0.920
0.292	0.049	0.968
0.149	0.025	0.993
0.041	0.007	1.000

The percentage of variation explained by each component is given in Table 10.9. The first two components explain almost 79% of the variance. The interpretation of the factors (the axes of Figure 10.16) is given in the table of correlations (Table 10.10). The first two columns of this table are plotted in Figure 10.17.

From Figure 10.17 (and Table 10.10) it appears that the first factor is a "size effect", it is positively correlated with all the variables describing the size of the activity of the companies. It is also a measure of the economic strength of the firms. The second factor describes the "shape" of the companies ("profit-cash flow" vs. "assets-sales" factor), which is more difficult to interpret from an economic point of view.

Example 10.10 Volle (1985) analyzes data on 28 individuals (Table B.14). For each individual, the time spent (in hours) on 10 different activities has been recorded over

Table 10.10 Correlations with PCs

	$r_{X_i Z_1}$	$r_{X_i Z_2}$	$r^2_{X_i Z_1} + r^2_{X_i Z_2}$
X_1: assets	0.47	−0.510	0.48
X_2: sales	0.78	−0.500	0.87
X_3: market value	0.89	−0.003	0.80
X_4: profits	0.59	0.770	0.95
X_5: cash flow	0.79	0.560	0.94
X_6: employees	0.76	−0.340	0.70

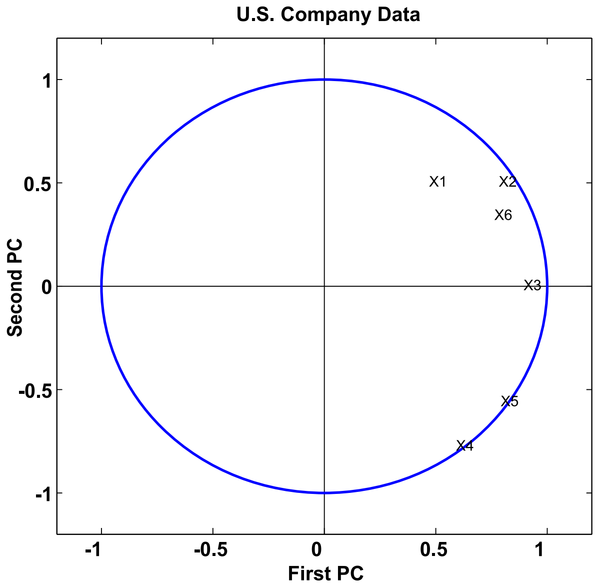


Fig. 10.17 The correlation of the original variables with the PCs `MVAnpcausco2i`

100 days, as well as informative statistics such as the individual's sex, country of residence, professional activity and matrimonial status. The results of a NPCA are given below.

The eigenvalues of the correlation matrix are given in Table 10.11. Note that the last eigenvalue is exactly zero since the correlation matrix is singular (the sum of all the variables is always equal to $2400 = 24 \times 100$). The results of the 4 first PCs are given in Tables 10.12 and 10.13.

From these tables (and Figures 10.18 and 10.19), it appears that the professional and household activities are strongly contrasted in the first factor. Indeed on the horizontal axis of Figure 10.18 it can be seen that all the active men are on the right and all the inactive women are on the left. Active women and/or single women are inbetween. The second factor contrasts meal/sleeping vs. toilet/shopping (note the high correlation between meal and sleeping). Along the vertical axis of Figure 10.18 we see near the bottom of the graph the people from Western-European countries,

Table 10.11 Eigenvalues of correlation matrix for the time budget data

ℓ_j	Proportion of variance	Cumulated proportion
4.59	0.459	0.460
2.12	0.212	0.670
1.32	0.132	0.800
1.20	0.120	0.920
0.47	0.047	0.970
0.20	0.020	0.990
0.05	0.005	0.990
0.04	0.004	0.999
0.02	0.002	1.000
0.00	0.000	1.000

Table 10.12 Correlation of variables with PCs

	$r_{X_i W_1}$	$r_{X_i W_2}$	$r_{X_i W_3}$	$r_{X_i W_4}$
X_1: prof	0.9772	−0.1210	−0.0846	0.0669
X_2: tran	0.9798	0.0581	−0.0084	0.4555
X_3: hous	−0.8999	0.0227	0.3624	0.2142
X_4: kids	−0.8721	0.1786	0.0837	0.2944
X_5: shop	−0.5636	0.7606	−0.0046	−0.1210
X_6: pers	−0.0795	0.8181	−0.3022	−0.0636
X_7: eati	−0.5883	−0.6694	−0.4263	0.0141
X_8: slee	−0.6442	−0.5693	−0.1908	−0.3125
X_9: tele	−0.0994	0.1931	−0.9300	0.1512
X_{10}: leis	−0.0922	0.1103	0.0302	−0.9574

who spend more time on meals and sleeping than people from the U. S. (who can be found close to the top of the graph). The other categories are inbetween.

In Figure 10.19 the variables television and other leisure activities hardly play any role (look at Table 10.12). The variable television appears in Z_3 (negatively correlated). Table 10.13 shows that this factor contrasts people from Eastern countries and Yugoslavia with men living in the U.S. The variable other leisure activities is the factor Z_4. It merely distinguishes between men and women in Eastern countries and in Yugoslavia. These last two factors are orthogonal to the preceeding axes and of course their contribution to the total variation is less important.

10.10 Exercises

Exercise 10.1 Prove Theorem 10.1. (Hint: use (4.23).)

Table 10.13 PCs for time budget data

	Z_1	Z_2	Z_3	Z_4
maus	0.0633	0.0245	−0.0668	0.0205
waus	0.0061	0.0791	−0.0236	0.0156
wnus	−0.1448	0.0813	−0.0379	−0.0186
mmus	0.0635	0.0105	−0.0673	0.0262
wmus	−0.0934	0.0816	−0.0285	0.0038
msus	0.0537	0.0676	−0.0487	−0.0279
wsus	0.0166	0.1016	−0.0463	−0.0053
mawe	0.0420	−0.0846	−0.0399	−0.0016
wawe	−0.0111	−0.0534	−0.0097	0.0337
wnwe	−0.1544	−0.0583	−0.0318	−0.0051
mmwe	0.0402	−0.0880	−0.0459	0.0054
wmwe	−0.1118	−0.0710	−0.0210	0.0262
mswe	0.0489	−0.0919	−0.0188	−0.0365
wswe	−0.0393	−0.0591	−0.0194	−0.0534
mayo	0.0772	−0.0086	0.0253	−0.0085
wayo	0.0359	0.0064	0.0577	0.0762
wnyo	−0.1263	−0.0135	0.0584	−0.0189
mmyo	0.0793	−0.0076	0.0173	−0.0039
wmyo	−0.0550	−0.0077	0.0579	0.0416
msyo	0.0763	0.0207	0.0575	−0.0778
wsyo	0.0120	0.0149	0.0532	−0.0366
maes	0.0767	−0.0025	0.0047	0.0115
waes	0.0353	0.0209	0.0488	0.0729
wnes	−0.1399	0.0016	0.0240	−0.0348
mmes	0.0742	−0.0061	−0.0152	0.0283
wmes	−0.0175	0.0073	0.0429	0.0719
mses	0.0903	0.0052	0.0379	−0.0701
fses	0.0020	0.0287	0.0358	−0.0346

Exercise 10.2 Interpret the results of the PCA of the U.S. companies. Use the analysis of the bank notes in Section 10.3 as a guide. Compare your results with those in Example 10.9.

Exercise 10.3 Test the hypothesis that the proportion of variance explained by the first two PCs for the U.S. companies is $\psi = 0.75$.

Exercise 10.4 Apply the PCA to the car data (Table B.7). Interpret the first two PCs. Would it be necessary to look at the third PC?

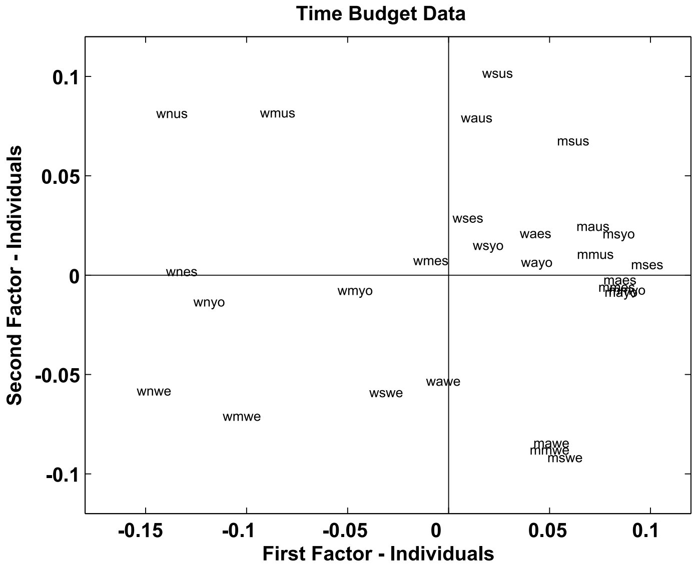


Fig. 10.18 Representation of the individuals MVAnpcatime

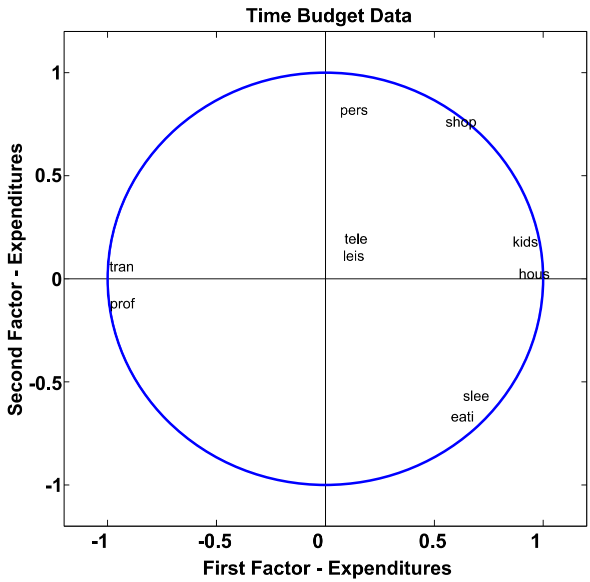


Fig. 10.19 Representation of the variables MVAnpcatime

Exercise 10.5 Take the athletic records for 55 countries (Appendix B.18) and apply the NPCA. Interpret your results.

Exercise 10.6 Apply a PCA to $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, where $\rho > 0$. Now change the scale of X_1, i.e., consider the covariance of cX_1 and X_2. How do the PC directions change with the screeplot?

Exercise 10.7 Suppose that we have standardized some data using the Mahalanobis transformation. Would it be reasonable to apply a PCA?

Exercise 10.8 Apply a NPCA to the U.S. CRIME data set (Table B.10). Interpret the results. Would it be necessary to look at the third PC? Can you see any difference between the four regions? Redo the analysis excluding the variable "area of the state."

Exercise 10.9 Repeat Exercise 10.8 using the U.S. HEALTH data set (Table B.16).

Exercise 10.10 Do a NPCA on the GEOPOL data set (see Table B.15) which compares 41 countries w.r.t. different aspects of their development. Why or why not would a PCA be reasonable here?

Exercise 10.11 Let U be an uniform r.v. on $[0, 1]$. Let $a \in \mathbb{R}^3$ be a vector of constants. Suppose that $X = Ua^\top = (X_1, X_2, X_3)$. What do you expect the NPCs of X to be?

Exercise 10.12 Let U_1 and U_2 be two independent uniform random variables on $[0, 1]$. Suppose that $X = (X_1, X_2, X_3, X_4)^\top$ where $X_1 = U_1$, $X_2 = U_2$, $X_3 = U_1 + U_2$ and $X_4 = U_1 - U_2$. Compute the correlation matrix $\mathcal{P}$ of X. How many PCs are of interest? Show that $\gamma_1 = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 1, 0)^\top$ and $\gamma_2 = (\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}, 0, 1)^\top$ are eigenvectors of $\mathcal{P}$ corresponding to the non trivial λ's. Interpret the first two NPCs obtained.

Exercise 10.13 Simulate a sample of size $n = 50$ for the r.v. X in Exercise 10.12 and analyze the results of a NPCA.

Exercise 10.14 Bouroche and Saporta (1980) reported the data on the state expenses of France from the period 1872 to 1971 (24 selected years) by noting the percentage of 11 categories of expenses. Do a NPCA of this data set. Do the three main periods (before WWI, between WWI and WWII, and after WWII) indicate a change in behavior w.r.t. to state expenses?