ST412-MULTIVARIATE STATISTICS WITH ADVANCED TOPICS
Assignment 1

Christis Katsouris 1154579
Department of Statistics
University of Warwick

December 5, 2011

1 Exercise 1

Let $\mathbf{X}$ be a $(n \times p)$ observed data matrix with variance/covariance matrix $\boldsymbol{\Sigma}$.

(i) The main properties of $\boldsymbol{\Sigma}$ as a matrix are as follow:

Definition 1. *The $(i,j)^{th}$ element of $\boldsymbol{\Sigma}$ matrix is given by*

$$\boldsymbol{\Sigma}_{ij} = \mathbb{E}[(\mathbf{X}_i - \boldsymbol{\mu}_i)(\mathbf{X}_j - \boldsymbol{\mu}_j)] = \boldsymbol{\Sigma}_{ij}$$

Properties:

1. From the above definition we see that for $i = j$ we get the variance of the $(i)^{th}$ variable. Hence the diagonal entries of the covariance matrix give the variances, i.e $(\boldsymbol{\Sigma})_{ij} = \mathbb{E}[(\mathbf{X}_i - \boldsymbol{\mu}_i)^2] = \sigma_i^2$

2. The covariance matrix is symmetric i.e $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^T$ since $(\boldsymbol{\Sigma})_{ij} = \sigma_{ij} = \sigma_{ji} = (\boldsymbol{\Sigma})_{ji}$

3. The covariance matrix is positive semi–definite.

   Proof. For $\underline{a} \in \mathbb{R}^n$ we have:

$$\begin{aligned} \mathbb{E}[(\mathbf{a}.(\mathbf{X} - \boldsymbol{\mu})^T)]^2 &= \mathbb{E}[\{\mathbf{a}.(\mathbf{X} - \boldsymbol{\mu})^T\}\{\mathbf{a}.(\mathbf{X} - \boldsymbol{\mu})^T\}^T] \geq 0 \\ \text{now using the property of transpose for matrices } (\mathbf{AB})^T &= \mathbf{B}^T\mathbf{A}^T \\ \implies \mathbb{E}[\{\mathbf{a}.(\mathbf{X} - \boldsymbol{\mu})T\}\{(\mathbf{X} - \boldsymbol{\mu}).\mathbf{a}^T\}] &\geq 0 \\ \implies \mathbf{a}.\boldsymbol{\Sigma}.\mathbf{a}^T &\geq 0 \end{aligned}$$

   □

4. The covariance matrix between linear transformation is given by: Let $\mathbf{a}$ and $\mathbf{b}$ be constant vectors $(p \times 1)$ and $\mathbf{X}$ be a random vector $(n \times p)$, then

$$Cov[\mathbf{aX}, \mathbf{bX}] = \mathbf{a}.Var[\mathbf{X}].\mathbf{b}^T$$

   Proof.

$$\begin{aligned} Cov[\mathbf{aX}, \mathbf{bX}] &= \mathbb{E}[(\mathbf{aX} - \mathbb{E}[\mathbf{aX}]).(\mathbf{bX} - \mathbb{E}[\mathbf{bX}])^T] \\ &= \mathbf{a}\mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}]).(\mathbf{X} - \mathbb{E}[\mathbf{X}])^T]\mathbf{b}^T \\ &= \mathbf{a}.Var[\mathbf{X}].\mathbf{b}^T \end{aligned}$$

   □

(ii) By saying that $\boldsymbol{\Sigma}$ is not scale-invariant we mean that if we re-scale one of the components of the matrix then we obtain different eigenvalues and eigenvectors. In other words it depends on the measure units of the data. For example if we have the covariance matrix $\boldsymbol{\Sigma}$ of some stock prices measured in Euro and we want to convert them into pounds the covariance matrix and anything computed based on the covariance matrix will change.

(iii) Let $\mathbf{X}_{n\times p}$ be the data matrix.

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$$

We can link the vector of means with the data matrix as follow:
First we consider the vector of means which is a column vector that has entries the means of each column of the data matrix $\mathbf{X}$. For example,

$$\bar{x}_1 = (x_{11}.1 + x_{21}.1 + \cdots + x_{n1}.1)/n$$

it gives the mean of the 1^{st} column of the data data matrix $\mathbf{X}$. Hence,

$$\bar{\mathbf{x}}_i = (x_{1i}.1 + x_{2i}.1 + \cdots + x_{ni}.1)/n$$

is the vector of the means of each column. Therefore, we have:

$$\bar{\mathbf{x}} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \ddots \\ \bar{x}_p \end{pmatrix} = \frac{1}{n} \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

Note also that the transpose matrix of the data matrix $\mathbf{X}$ is a $(p \times n)$ matrix, i.e

$$\mathbf{X}^T = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{pmatrix}$$

Hence,

$$\bar{\mathbf{x}} \;=\; \frac{1}{n}\mathbf{X}^T_{n\times p}\mathbf{1}_{p\times 1} \tag{1}$$

Now, taking the transpose of (1) and pre–multiplying on both sides by the column vector of 1's, we get:

$$\mathbf{1}\bar{\mathbf{x}}^T \;=\; \frac{1}{n}\mathbf{1}\mathbf{1}^T\mathbf{X} = \begin{pmatrix} \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_p \\ \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_p \\ \vdots & \vdots & \ddots & \vdots \\ \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_p \end{pmatrix} \tag{2}$$

We can produce the residual matrix by substructing (2) from $\mathbf{X}$:

$$\mathbf{X} - \frac{1}{n}\mathbf{1}\mathbf{1}^T\mathbf{X} = \begin{pmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & \cdots & x_{1p} - \bar{x}_p \\ x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 & \cdots & x_{2p} - \bar{x}_p \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} - \bar{x}_1 & x_{n2} - \bar{x}_2 & \cdots & x_{np} - \bar{x}_p \end{pmatrix} \tag{3}$$

We should note that,

$$\begin{aligned} &(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T)^T(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T) \\ &\qquad = (\mathbf{I} - \frac{1}{n}\mathbf{1}^T\mathbf{1})(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T) \\ &\qquad = (\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T - \frac{1}{n}\mathbf{1}^T\mathbf{1} + \frac{1}{n^2}\mathbf{1}^T\mathbf{1}\mathbf{1}\mathbf{1}^T) \\ &\qquad = (\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T - \frac{1}{n}\mathbf{1}^T\mathbf{1} + \frac{1}{n}\mathbf{1}^T\mathbf{1}) \\ &\qquad = (\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T) \end{aligned}$$

$$(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T)^T(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T) = (\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T) \tag{4}$$

Note:

1. $\mathbf{I}$ is an orthogonal matrix i.e $\mathbf{I}.\mathbf{I}^T = \mathbf{I}^T.\mathbf{I} = \mathbf{I}$.
2. For matrices $\mathbf{A}$ and $\mathbf{B}$, $(\mathbf{AB})^T = (\mathbf{B}^T\mathbf{A}^T)$ and $(\mathbf{A}^T)^T = \mathbf{A}$
3. The vector multiplication $\mathbf{1}\mathbf{1}^T = \sum_{i=1}^{n} 1 = n$

Now the matrix $n\mathbf{\Sigma}$ is the matrix that has the sum of squares, i.e the square difference from the mean. This can be obtained by the multiplication of matrix given by (3) and its transpose.

$$n\mathbf{\Sigma} = (\mathbf{X} - \frac{1}{n}\mathbf{1}\mathbf{1}^T)^T(\mathbf{X} - \frac{1}{n}\mathbf{1}\mathbf{1}^T)$$

$$n\mathbf{\Sigma} = \begin{pmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & \cdots & x_{1p} - \bar{x}_p \\ x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 & \cdots & x_{2p} - \bar{x}_p \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} - \bar{x}_1 & x_{n2} - \bar{x}_2 & \cdots & x_{np} - \bar{x}_p \end{pmatrix} \begin{pmatrix} x_{11} - \bar{x}_1 & x_{21} - \bar{x}_2 & \cdots & x_{n1} - \bar{x}_p \\ x_{12} - \bar{x}_1 & x_{22} - \bar{x}_2 & \cdots & x_{n2} - \bar{x}_p \\ \vdots & \vdots & \ddots & \vdots \\ x_{1p} - \bar{x}_1 & x_{2p} - \bar{x}_2 & \cdots & x_{pn} - \bar{x}_p \end{pmatrix} \tag{5}$$

Now pre–multiplying both sides of (4) by $\mathbf{X}^T$ gives:

$$\mathbf{X}^T(\mathbf{I}-\frac{1}{n}\mathbf{1}\mathbf{1}^T)^T(\mathbf{I}-\frac{1}{n}\mathbf{1}\mathbf{1}^T)=\mathbf{X}^T(\mathbf{I}-\frac{1}{n}\mathbf{1}\mathbf{1}^T)$$

$$(\mathbf{X}-\frac{1}{n}\mathbf{1}\mathbf{1}^T\mathbf{X})^T(\mathbf{I}-\frac{1}{n}\mathbf{1}\mathbf{1}^T)=(\mathbf{X}^T-\frac{1}{n}\mathbf{X}^T\mathbf{1}\mathbf{1}^T)$$

and post–multiplying both sides of the above equation with $\mathbf{X}$ gives:

$$(\mathbf{X}-\frac{1}{n}\mathbf{1}\mathbf{1}^T\mathbf{X})^T(\mathbf{X}-\frac{1}{n}\mathbf{1}\mathbf{1}^T)=(\mathbf{X}^T\mathbf{X}-\frac{1}{n}\mathbf{X}^T\mathbf{1}\mathbf{1}^T\mathbf{X}) \tag{6}$$

But the left hand side of (6) is just $n\mathbf{\Sigma}$. Hence we can conclude that

$$\mathbf{\Sigma}=\frac{1}{n}(\mathbf{X}^T\mathbf{X}-\frac{1}{n}\mathbf{X}^T\mathbf{1}\mathbf{1}^T\mathbf{X})$$

2 Exercise 2

We consider $\mathbf{X}$ a vector with bivariate normal distribution given by:

$$\mathbf{X}=\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$$

We need to show that $\mathbf{a}^T\mathbf{X}$ and $\mathbf{b}^T\mathbf{X}$ are independent where $\mathbf{a}=\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\mathbf{b}=\begin{pmatrix} 1 \\ -1 \end{pmatrix}$

Therefore we consider a matrix $\mathbf{A}$ that has as columns the column vectors $\mathbf{a}$ and $\mathbf{b}$ given by:

$$\mathbf{A}=(\mathbf{a},\mathbf{b})=\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

Now using properties of the expectation and the variance/covariance matrix of the multivariate normal distribution we can find the mean vector and the variance/covariance matrix of the joint distribution of these two vectors.

$$\mathbb{E}[\mathbf{A}.\mathbf{X}]=\mathbf{A}.\mathbb{E}[\mathbf{X}]=\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\times\begin{pmatrix} 2 \\ 2 \end{pmatrix}=\begin{pmatrix} 4 \\ 0 \end{pmatrix}$$

because of the linearity property of the expectation and

$$Var[\mathbf{A}.\mathbf{X}] = \mathbf{A}.Var[\mathbf{X}].\mathbf{A}^T = \mathbf{A}.\mathbf{\Sigma}.\mathbf{A}^T = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$
$$= \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

Hence,

$$\mathbf{AX} \sim \mathcal{N}\left(\begin{bmatrix} 4 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \right)$$

As we can see from the above variance/covariance matrix of the joint distribution, the correlation between the two random variables is zero.
i.e.

$$Cov(\mathbf{a}^T\mathbf{X}, \mathbf{b}^T\mathbf{X}) = Cov(\mathbf{b}^T\mathbf{X}, \mathbf{a}^T\mathbf{X}) = 0$$

Furthermore we know that they are normally distributed so we can conclude that $\mathbf{a}^T\mathbf{X}$ and $\mathbf{b}^T\mathbf{X}$ are independent.
Finally their marginal probability distributions are:

$$\mathbf{a}^T\mathbf{X} \sim \mathcal{N}(4, 2) \text{ and } \mathbf{b}^T\mathbf{X} \sim \mathcal{N}(0, 2)$$

3 Exercise 3

In this exercise we have the vectors $\mathbf{X}$ and $\mathbf{Y}|\mathbf{X}$ to follow multivariate normal distribution given as below:

$$\mathbf{X} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \right)$$

and

$$\mathbf{Y}|\mathbf{X} \sim \mathcal{N}\left(\begin{bmatrix} X_1 \\ X_1 + X_2 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right)$$

We need to determine the distribution of $\mathbf{W} = \mathbf{Y} - \mathbf{X}$
Therefore we can use the following theorem, that has been used in lectures. Also it can be found in the book "Applied Multivariate Statistical Analysis" by $W.Hardle$ & $L.Simar$, pages 149–150. [2]

Theorem 1. *If* $\mathbf{X_1} \sim \mathcal{N}_r(\mu_1, \mathbf{\Sigma}_{11})$ *and* $(\mathbf{X_2}|\mathbf{X_1} = \mathbf{x_1}) \sim \mathcal{N}_{p-r}(\mathbf{A}\mathbf{x_1} + \mathbf{b}, \mathbf{B})$ *where* $\mathbf{B}$ *does not depend on* $\mathbf{x_1}$*, then* $\mathbf{X} = \begin{pmatrix} \mathbf{X_1} \\ \mathbf{X_2} \end{pmatrix} \sim \mathcal{N}_p(\mu, \mathbf{\Sigma})$ *,where*

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu_1} \\ \mathbf{A}\mu_\mathbf{1} + \mathbf{b} \end{pmatrix} \text{ and } \mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma_{11}} & \mathbf{\Sigma_{11}A^T} \\ \mathbf{\Sigma_{11}A} & \mathbf{B} + \mathbf{A\Sigma_{11}A^T} \end{pmatrix}.$$

Using the above theorem we set $\boldsymbol{\mu}_1 = \begin{pmatrix}1\\2\end{pmatrix}$, $\boldsymbol{\Sigma}_{11} = \begin{pmatrix}2 & 1\\1 & 2\end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}$.

Let $\mathbf{A}\mathbf{x_1} = \begin{pmatrix}1 & 0\\1 & 1\end{pmatrix}\begin{pmatrix}X_1\\X_2\end{pmatrix} = \begin{pmatrix}X_1\\X_1+X_2\end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix}0\\0\end{pmatrix}$

so $\mathbf{A}\mu_1 + \mathbf{b} = \begin{pmatrix}1 & 0\\1 & 1\end{pmatrix}\begin{pmatrix}1\\2\end{pmatrix} + \begin{pmatrix}0\\0\end{pmatrix} = \begin{pmatrix}1\\3\end{pmatrix}$

Hence $\boldsymbol{\mu} = \begin{pmatrix}\mu_X\\\mu_Y\end{pmatrix} = \begin{pmatrix}1 & 2 & 1 & 3\end{pmatrix}^T$. Now we need to compute the covariance matrix. We have:

$$\mathbf{A\Sigma_{11}} = \begin{pmatrix}1 & 0\\1 & 1\end{pmatrix}\begin{pmatrix}2 & 1\\1 & 2\end{pmatrix} = \begin{pmatrix}2 & 1\\3 & 3\end{pmatrix}, \mathbf{\Sigma_{11}A}^T = \begin{pmatrix}2 & 3\\1 & 3\end{pmatrix}\begin{pmatrix}1 & 0\\1 & 1\end{pmatrix} = \begin{pmatrix}2 & 1\\3 & 3\end{pmatrix}.$$

so

$$\mathbf{B} + \mathbf{A\Sigma_{11}A}^T = \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix} + \begin{pmatrix}2 & 1\\3 & 3\end{pmatrix}\begin{pmatrix}1 & 0\\1 & 1\end{pmatrix} = \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix} + \begin{pmatrix}2 & 3\\3 & 6\end{pmatrix} = \begin{pmatrix}3 & 3\\3 & 7\end{pmatrix}$$

i.e

$$\boldsymbol{\Sigma} = \begin{array}{c c} & \begin{array}{cccc} X_1 & X_2 & Y_1 & Y_2 \end{array} \\ \begin{array}{c} X_1 \\ X_2 \\ Y_1 \\ Y_2 \end{array} & \begin{pmatrix} 2 & 1 & 2 & 3 \\ 1 & 2 & 1 & 3 \\ 2 & 1 & 3 & 3 \\ 3 & 3 & 3 & 7 \end{pmatrix} \end{array}$$

Therefore,

$$\mathbf{Y} \sim \mathcal{N}\left(\begin{bmatrix}1\\3\end{bmatrix}, \begin{bmatrix}3 & 3\\3 & 7\end{bmatrix}\right)$$

.

Since we have the distribution of $\mathbf{Y}$ we can now compute the distribution of $\mathbf{W} = \mathbf{Y} - \mathbf{X}$.

$\mathbb{E}[\mathbf{W}] = \mathbb{E}[\mathbf{Y} - \mathbf{X}] = \mathbb{E}[\mathbf{Y}] - \mathbb{E}[\mathbf{X}] = \begin{pmatrix}1\\3\end{pmatrix} - \begin{pmatrix}1\\2\end{pmatrix} = \begin{pmatrix}0\\1\end{pmatrix}$

We compute the covariance matrix as follow:

$$\mathbf{W} = \mathbf{Y} - \mathbf{X} = \begin{pmatrix}Y_1 - X_1\\Y_2 - X_2\end{pmatrix} = \begin{pmatrix}-1 & 0 & 1 & 0\\0 & -1 & 0 & 1\end{pmatrix}\begin{pmatrix}X_1\\X_2\\Y_1\\Y_2\end{pmatrix}$$

Let $\mathbf{Z} = \begin{pmatrix}\mathbf{X}\\\mathbf{Y}\end{pmatrix}$ and $\mathbf{A} = \begin{pmatrix}-1 & 0 & 1 & 0\\0 & -1 & 0 & 1\end{pmatrix}$ then

$Var[\mathbf{A}.\mathbf{Z}] = \mathbf{A}.Var[\mathbf{Z}].\mathbf{A}^T = \mathbf{A}.\boldsymbol{\Sigma}.\mathbf{A}^T$

$$= \begin{pmatrix} -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 & 2 & 3 \\ 1 & 2 & 1 & 3 \\ 2 & 1 & 3 & 3 \\ 3 & 3 & 3 & 7 \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 0 & -1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 2 & 1 & 2 & 4 \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 0 & -1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$= \begin{array}{c} \\ W_1 \\ W_2 \end{array} \begin{array}{c} \begin{array}{cc} W_1 & W_2 \end{array} \\ \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} \end{array}$$

Note that we can also check this, by proceeding as follow:

$$Var(W_1) = Var(Y_1 - X_1) = Var(Y_1) - Var(X_1) - Cov(Y_1, X_1) = 3 + 2 - 4 = 1$$
$$Var(W_2) = Var(Y_2 - X_2) = Var(Y_2) - Var(X_2) - Cov(Y_2, X_2) = 7 + 2 - 6 = 3$$

$$\begin{aligned} Cov(W_1, W_2) &= Cov(Y_1 - X_1, Y_2 - X_2) \\ &= Cov(Y_1, Y_2) - Cov(Y_1, X_2) - Cov(X_1, Y_2) - Cov(X_1, X_2) \\ &= 3 - 1 - 3 + 1 = 0 \end{aligned}$$

Summarising the above:

$$\mathbf{W} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \right)$$

4 Exercise 4

(a) The covariance matrix $\mathbf{\Sigma}$ of the given dataset is found to have ranked eigenvalues

$$\boldsymbol{\lambda} = \begin{pmatrix} 5.56 & 1.15 & 0.37 & 0.1 & 0.08 & 0.05 & 0.04 & 0.02 \end{pmatrix}^T$$

with corresponding vector $\boldsymbol{\psi_j}$ of cumulative proportion of explained variance equal to

$$\boldsymbol{\psi} = \begin{pmatrix} 0.76 & 0.91 & 0.96 & 0.99 & 0.99 & 1.00 & 1.00 \end{pmatrix}^T$$

That is the j-th component of $\boldsymbol{\psi}$ is defined by

$$\boldsymbol{\psi} = \frac{\sum_{i=1}^{j} \lambda_i}{det(\mathbf{\Sigma})} \text{ where } j = 1, ...7$$

The idea of principal components is by starting with the variables $X_1, ...X_p$ to find a transformation of the data i.e $Y_1, ...Y_p$ (principal components) so that these new variables are uncorrelated with the further property that:

$$Var(Y_1) \geq Var(Y_2) \geq .... \geq Var(Y_p)$$

So Y_1 is more important that Y_1. Thus, we are interested in the first few principal components, since our aim is to reduce the dimensions of the data. That is to summarize the data with a smaller number of variables but at the same time losing as little information as possible.
From the vector $\boldsymbol{\psi_j}$ of cumulative proportion of explained variance, we see that the first PC explains the 76% of the variability of the data and the second PC explains the 91% of the variability of the data. Therefore we can consider just the first two principal components as an adequate number of principal components that explains the variability of the data.

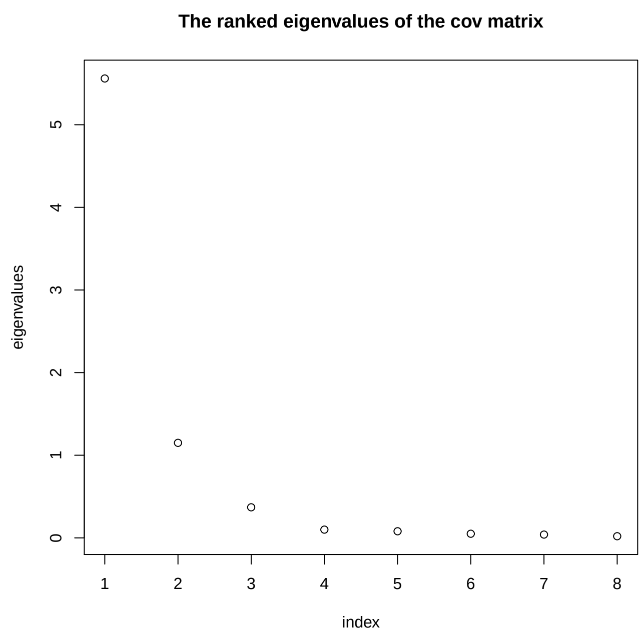

Figure 1: Scree plot

Another way to check how many principal components are appropriate to be used is to look at a scree plot. (*See reference 1 pages 444-445*)As we can see from figure 1 that shows the scree plot, the eigenvalues after λ_3 they are relatively small and they have approximately the same size. Therefore we can conclude that two principal components are enough to explain the variability of the data.

(b) The given figure (the circle) is based on the idea of the following problem. Consider we are interested to calculate the covariance between the data and the principal components. Let $\mathbf{X}_i$ represents the data and $\mathbf{y}_i$ the principal components. Then

$$\begin{aligned} Cov(\mathbf{X}, \mathbf{y}) = Cov(\mathbf{X}, \mathbf{\Gamma}^T(\mathbf{X} - \boldsymbol{\mu})) = Cov(\mathbf{X}, \mathbf{X})\mathbf{\Gamma} &= \mathbf{\Sigma\Gamma} \\ &= (\mathbf{\Gamma\Lambda\Gamma}^T)\mathbf{\Gamma} = \mathbf{\Gamma\Lambda} \\ \text{using the spectral decomposition of } \mathbf{\Sigma} \end{aligned}$$

Hence,

$$\rho_{ij} = Corr(X_i, y_j) = \frac{\gamma_{ij}\lambda_j}{\sqrt{\sigma_{ii}\lambda_j}} = \gamma_{ij}\sqrt{\frac{\lambda_j}{\sigma_{ii}}}$$

and note that

$$\sum_{i=1}^{p} \rho_{ij}^2 = \sum_{i=1}^{p} Corr^2(X_i, y_j) = 1$$

Therefore we consider the graph of the 1st PC and 2nd PC, which is basically the graph of ρ_{ij} versus ρ_{ji}. The plot shows which of the original random variables are correlated with the $PC1$ and the $PC2$. i.e we can see how well they explained the data.

Note that all the variables should lie within the circle. In fact if the variables are close to the periphery of the circle then that is a good evidence that they are explained well by the PCs, which is the case here. On the other hand though, if we look at the $PC1$ coordinate of the $(PC1, PC2)$ plane, we observe that most of the variables have $PC1$ coordinate greater then 0.8 which means that they more correlated to the first principal component. Since $PC1$ and $PC2$ are independent then it follows that the variables can't be highly correlated with the second principal component. For instance the *sport* variable has $PC2$ coordinate around 0.3 so it can't be correlated with the $PC1$.

More specifically, we see that $PC1$ explains for example the variables *price* and *value* , since they are above the horizontal axis. These variables have high correlation, meaning that someone needs to pay more in order to get a car with a better design. Moreover we can say that $PC2$ describes for example the difference between *easy* and the sum of *price* and *value* since the 1st one lie above the horizontal axis and the other two below. In other words how easy handling is a car doesn't depend on the price of the car.

The above analysis suggests that the variables have a significantly stronger correlation with the first principal component.

5 Exercise 5

Let U_1 and U_2 be two independent and identically distributed *iid* random variables with common distribution Uniform on [0,1]. We consider the random vector $\mathbf{X} = \left(X_1 X_2 X_3 X_4\right)^T$ where $X_1 = U_1, X_2 = U_2, X_3 = U_1 + U_2, X_4 = U_1 - U_2$.

(a) We compute:

(i) The mean vector $\boldsymbol{\mu} = \mathbb{E}[\mathbf{X}]$

Remark 1. *If* $X \sim Unif([a, b])$ *then* $\mathbb{E}[X] = \frac{1}{2}(a + b)$ *and* $Var(X) = \frac{1}{12}(b - a)^2$.

Therefore,

$$\begin{aligned}
\mathbb{E}[X_1] = \mathbb{E}[U_1] &= \frac{1}{2} \\
\mathbb{E}[X_2] = \mathbb{E}[U_2] &= \frac{1}{2} \\
\mathbb{E}[X_3] = \mathbb{E}[U_1 + U_2] = \mathbb{E}[U_1] + \mathbb{E}[U_2] &= 1 \\
\mathbb{E}[X_4] = \mathbb{E}[U_1 - U_2] = \mathbb{E}[U_1] - \mathbb{E}[U_1] &= 0
\end{aligned}$$

so $\boldsymbol{\mu} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 1 & 0 \end{pmatrix}^T$.

(ii) The covariance matrix $\mathbf{\Sigma}$ of $\mathbf{X}$:

We have:

$$\begin{aligned}
Var[X_1] = Var[U_1] = \frac{1}{12} \text{ and } Var[X_2] = Var[U_2] &= \frac{1}{12} \\
Var[X_3] = Var[U_1 + U_2] = Var[U_1] + Var[U_2] = \frac{1}{12} + \frac{1}{12} &= \frac{1}{6} \\
Var[X_4] = Var[U_1 - U_2] = Var[U_1] + Var[U_2] = \frac{1}{12} + \frac{1}{12} &= \frac{1}{6}
\end{aligned}$$

Now we need to compute the covariance between the variables $X_1 ... X_4$. We see that X_1 and X_2 are independent random variables, but not all of them are pairwise independent.

$$Cov(X_1, X_2) \quad = Cov(X_2, X_1) = 0$$

$$\begin{aligned}
Cov(X_1, X_3) &= Cov(U_1, U_1 + U_2) \\
&= Cov(U_1, U_1) + Cov(U_1, U_2) \\
&= Var(U_1) + 0 = \frac{1}{12} \\
&= Cov(X_3, X_1)
\end{aligned}$$

$$\begin{aligned}
Cov(X_1, X_4) &= Cov(U_1, U_1 - U_2) \\
&= Cov(U_1, U_1) - Cov(U_1, U_2) \\
&= Var(U_1) + 0 = \frac{1}{12} \\
&= Cov(X_4, X_1)
\end{aligned}$$

$$\begin{aligned}
Cov(X_2, X_4) &= Cov(U_2, U_1 - U_2) \\
&= Cov(U_2, U_1) - Cov(U_2, U_2) \\
&= 0 - Var(U_2) = -\frac{1}{12} \\
&= Cov(X_4, X_2)
\end{aligned}$$

Therefore the covariance matrix is given by:

$$\boldsymbol{\Sigma} = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \end{array} \begin{array}{c} \begin{array}{cccc} X_1 & X_2 & X_3 & X_4 \end{array} \\ \begin{pmatrix} \frac{1}{12} & 0 & \frac{1}{12} & \frac{1}{12} \\ 0 & \frac{1}{12} & \frac{1}{12} & -\frac{1}{12} \\ \frac{1}{12} & \frac{1}{12} & \frac{1}{6} & 0 \\ \frac{1}{12} & -\frac{1}{12} & 0 & \frac{1}{6} \end{pmatrix} \end{array}$$

(iii) The corelation matrix $\mathbf{R}$ of $\mathbf{X}$:

$$\rho_{11} = \frac{\sigma_{11}}{\sqrt{\sigma_{11}.\sigma_{11}}} = \frac{\frac{1}{12}}{\frac{1}{12}} = 1 = \rho_{22} = \rho_{33} = \rho_{44}$$

$$\rho_{12} = \frac{\sigma_{12}}{\sqrt{\sigma_{11}.\sigma_{22}}} = 0 = \rho_{21}$$

$$\rho_{13} = \frac{\sigma_{13}}{\sqrt{\sigma_{11}.\sigma_{33}}} = \frac{\frac{1}{12}}{\sqrt{\frac{1}{12}\frac{1}{6}}} = \frac{\frac{1}{12}}{\frac{6}{\sqrt{12}}} = \frac{\sqrt{2}}{2} = \rho_{31}$$

$$\rho_{14} = \frac{\sigma_{14}}{\sqrt{\sigma_{11}.\sigma_{44}}} = \frac{\frac{1}{12}}{\sqrt{\frac{1}{12}\frac{1}{6}}} = \frac{\frac{1}{12}}{\frac{6}{\sqrt{12}}} = \frac{\sqrt{2}}{2} = \rho_{41}$$

$$\rho_{23} = \frac{\sigma_{23}}{\sqrt{\sigma_{22}.\sigma_{33}}} = \frac{\frac{1}{12}}{\sqrt{\frac{1}{12}\frac{1}{6}}} = \frac{\frac{1}{12}}{\frac{6}{\sqrt{12}}} = \frac{\sqrt{2}}{2} = \rho_{32}$$

$$\rho_{24} = \frac{\sigma_{24}}{\sqrt{\sigma_{22}.\sigma_{44}}} = \frac{-\frac{1}{12}}{\sqrt{\frac{1}{12}\frac{1}{6}}} = -\frac{\frac{1}{12}}{\frac{6}{\sqrt{12}}} = -\frac{\sqrt{2}}{2} = \rho_{41}$$

Hence the correlation matrix is given by:

$$\mathbf{R} = \begin{pmatrix} 1 & 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ 0 & 1 & \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 1 & 0 \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 1 \end{pmatrix}$$

(b) We need to show that

$$\boldsymbol{\gamma}_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 1 & 0 \end{pmatrix}$$
$$\boldsymbol{\gamma}_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & 1 \end{pmatrix}$$

are two eigenvalues of the correlation matrix $\mathbf{R}$ associated with non-zero eigenvalues. Therefore we can simply prove that the eigenvalue-eigenvector equation holds, i.e $\mathbf{R}\boldsymbol{\gamma}_\mathbf{i} = \lambda\boldsymbol{\gamma}_i$, where i={ 1,2 }.

For eigenvector $\boldsymbol{\gamma}_1$ we have:

$$\mathbf{R}\gamma_\mathbf{1} = \begin{pmatrix} 1 & 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ 0 & 1 & \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 1 & 0 \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{2}{\sqrt{2}} \\ \frac{2}{\sqrt{2}} \\ 2 \\ 0 \end{pmatrix} = 2 \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 1 \\ 0 \end{pmatrix} = 2\boldsymbol{\gamma}_1$$

Hence $\lambda_1 = 2 > 0$

For eigenvector $\gamma_\mathbf{2}$ we have:

$$\mathbf{R}\boldsymbol{\gamma}_2 = \begin{pmatrix} 1 & 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ 0 & 1 & \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 1 & 0 \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{2}{\sqrt{2}} \\ \frac{-2}{\sqrt{2}} \\ 2 \\ 0 \end{pmatrix} = 2 \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 1 \\ 0 \end{pmatrix} = 2\boldsymbol{\gamma}_2$$

Hence $\lambda_2 = 2 > 0$

(c) Prove that the matrix $\mathbf{R}$ has two zero eigenvalues.

Proof. We already know that two eigenvalues of $\mathbf{R}$ are $\lambda_1 = \lambda_2 = 2$. We need to show that the other two are zero. Note that the covariance and correlation matrix are symmetric matrices therefore their eigenvalues and eigenvectors are real. Also they are square matrices. We know that the trace of a square matrix is just the sum of its eigenvalues and the determinant of a matrix is the product of its eigenvalues.

$$Tr(\mathbf{R}) = \sum_{i=1}^{4} \lambda_i \Longrightarrow \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 4$$

But

$$\lambda_1 + \lambda_2 = 4 \Longrightarrow \lambda_3 + \lambda_4 = 0$$

Also,

$$det(\mathbf{R}) = \prod_{i=1}^{4} \lambda_i \Longrightarrow \lambda_1.\lambda_2.\lambda_3.\lambda_4 = 0 \Longrightarrow 4\lambda_3.\lambda_4 = 0$$

Note that the matrix **R** doesn't have full rank, because its columns are not linearly independent. Therefore the determinant of this matrix is zero. (By a theorem from linear algebra)

Therefore $\lambda_3 = \lambda_4 = 0$ $\square$

(d) Based on $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$, write the formula describing the first two principal components of **X**.

Since $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ are the eigenvectors of the correlation matrix and we want the principal components of **X**, they are given by:

$$\mathbf{Y_i} = \boldsymbol{\gamma}_i \mathbf{Z_i}, i = 1, ...4 \ , \mathbf{Z_i} = \left(\frac{\mathbf{X_i} - \boldsymbol{\mu}_i}{\sqrt{Var(\mathbf{X_i})}} \right)^T$$

$$\begin{aligned}
Y_1 &= \begin{pmatrix} 1 & 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{pmatrix} \\
&= \frac{1}{\sqrt{2}} Z_1 + \frac{1}{\sqrt{2}} Z_2 + Z_3 \\
&= \frac{1}{\sqrt{2}} \frac{X_1 - \mu_1}{\sqrt{Var(X_1)}} + \frac{1}{\sqrt{2}} \frac{X_2 - \mu_2}{\sqrt{Var(X_2)}} + \frac{X_3 - \mu_3}{\sqrt{Var(X_3)}} \\
&= \frac{1}{\sqrt{2}} \frac{X_1 - \frac{1}{2}}{\frac{1}{\sqrt{2}}} + \frac{1}{\sqrt{2}} \frac{X_2 - \frac{1}{2}}{\frac{1}{\sqrt{6}}} + \frac{X_3 - 1}{\frac{1}{\sqrt{6}}} \\
&= (X_1 - \frac{1}{2}) + (X_2 - \frac{1}{2}) + \sqrt{6}(X_2 - 1) \\
&= X_1 + X_2 + \sqrt{6} X_3 - 1 - \sqrt{6}
\end{aligned}$$

$$
\begin{aligned}
Y_2 &= \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & 1 \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{pmatrix} \\
&= \frac{1}{\sqrt{2}} Z_1 - \frac{1}{\sqrt{2}} Z_2 + Z_4 \\
&= \frac{1}{\sqrt{2}} \frac{X_1 - \mu_1}{\sqrt{Var(X_1)}} - \frac{1}{\sqrt{2}} \frac{X_2 - \mu_2}{\sqrt{Var(X_2)}} + \frac{X_4 - \mu_4}{\sqrt{Var(X_4)}} \\
&= \frac{1}{\sqrt{2}} \frac{X_1 - \frac{1}{2}}{\frac{1}{\sqrt{2}}} - \frac{1}{\sqrt{2}} \frac{X_2 - \frac{1}{2}}{\frac{1}{\sqrt{6}}} + \frac{X_4 - 0}{1} \\
&= (X_1 - \frac{1}{2}) - (X_2 - \frac{1}{2}) + \sqrt{6} X_4 \\
&= X_1 + X_2 + \sqrt{6} X_4
\end{aligned}
$$

(e) Show with a numerical example that the principal components are not scale-invariant.

It is clear that the principal components are not scale–invariant i.e applying principal components analysis using the covariance matrix gives different results than doing the same using the correlation matrix. A numerical example of this would be to consider a covariance matrix (2×2), find its eigenvalues and compare these eigenvalues with the corresponding eigenvalues of the correlation matrix.

Let

$$\mathbf{\Sigma} = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}$$

Then we compute the eigenvalues of this covariance matrix:

$$
\begin{aligned}
det(\mathbf{\Sigma} - \lambda \mathbf{I}) = 0 &\Longrightarrow (\lambda - 3)(\lambda - 2) - 1 = 0 \\
&\Longrightarrow \lambda^2 - 5\lambda - 5 = 0 \\
&\Longrightarrow \lambda_1 = \frac{-5}{2} + \frac{\sqrt{5}}{2} \text{ and } \lambda_2 = \frac{-5}{2} - \frac{\sqrt{5}}{2}
\end{aligned}
$$

The correlation matrix is given by:

$$\mathbf{R} = \begin{pmatrix} 1 & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} & 1 \end{pmatrix}$$

and its eigenvalues are:

$$
\begin{aligned}
det(\mathbf{R} - \lambda^* \mathbf{I}) = 0 &\Longrightarrow (\lambda^* - 1)^2 - \frac{1}{6} = 0 \\
&\Longrightarrow (\lambda^* - 1)^2 = \frac{1}{6} \\
&\Longrightarrow \lambda_1^* = \frac{7}{6} \text{ and } \lambda_2^* = \frac{5}{6}
\end{aligned}
$$

As we can observe from this example the eigenvalues of a covariance and correlation matrix are different, meaning that their corresponding eigenvectors and principal components will be different.

(f) Let us suppose that we have operated the *Mahalanobis* transformation to our data, i.e $\mathbf{\Sigma}^{-1} = \mathbf{\Sigma}^{-\frac{1}{2}}.\mathbf{\Sigma}^{-\frac{1}{2}}$
Set

$$\mathbf{Z} = \mathbf{\Sigma}^{-\frac{1}{2}}(\mathbf{X} - \bar{\mathbf{x}}) \quad = \quad \mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{X} - \mathbf{\Sigma}^{-\frac{1}{2}}\bar{\mathbf{x}}$$

but this is a linear transformation of the form $\mathbf{Y} = \mathbf{AX} - \mathbf{b}$ and hence $Var(\mathbf{Y}) = \mathbf{A}Var(\mathbf{X})\mathbf{A}^T$
so $Var(\mathbf{Z}) = \mathbf{\Sigma}^{-\frac{1}{2}}.\mathbf{\Sigma}.\mathbf{\Sigma}^{-\frac{1}{2}} = \mathbf{I}$
Therefore the new covariance matrix after applying the *Mahalanobis* transformation to the data will be the identity matrix. Thus, the eigenvalues are non-distinct and so the eigenvectors will be orthogonal. We can apply this transformation.

References

[1] Richard A. Johnson & Dean W. Wichen, *Applied Multivariate Statistical Analysis*, Prentice Hall.

[2] W. Hardle & L. Simar, *Applied Multivariate Statistical Analysis*, Springer, 2nd Edition.