MVA_HW5

陈子睿 15220212202842

5/21/24

1 Multivariate Regression Model

(a) First, we have $\hat{\mathbf{B}} = \left(\hat{\beta}_{(1)}, \hat{\beta}_{(2)}, \ldots, \hat{\beta}_{(p)}\right) = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{B} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\Xi$, where $\Xi = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)'$ and

$\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are multinormally distributed with $\epsilon_i \sim N\left(0, \Sigma = (\sigma_{ij})_{n\times n}\right)$. Thus,

$$\begin{aligned} Cov\left(\hat{\beta}_{(i)}, \hat{\beta}_{(j)}\right) &= Cov\left(\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\hat{\boldsymbol{\epsilon}}_{(i)}, \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\hat{\boldsymbol{\epsilon}}_{(j)}\right) \\ &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\sigma_{ij}I_n\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} = \sigma_{ij}\left(\mathbf{X}'\mathbf{X}\right)^{-1}, \end{aligned}$$

$$Cov\left(\mathrm{x}_0'\hat{\beta}_{(i)}, \mathrm{x}_0'\hat{\beta}_{(j)}\right) = \sigma_{ij}\mathrm{x}_0'\left(\mathrm{X}'\mathbf{X}\right)^{-1}\mathrm{x}_0.$$

which leads to

$$Cov\left(\mathrm{x}_0'\hat{\beta}_{(i)}, \mathrm{x}_0'\hat{\beta}_{(j)}\right) = \sigma_{ij}\mathrm{x}_0'\left(\mathrm{X}'\mathbf{X}\right)^{-1}\mathrm{x}_0.$$

Now, we have

$$\hat{\mathrm{y}}_0 = \hat{\mathrm{B}}'\mathrm{x}_0 + \epsilon_0 = \left(\mathrm{x}_0'\hat{\beta}_{(1)}, \mathrm{x}_0'\hat{\beta}_{(2)}, \ldots, \mathrm{x}_0'\hat{\beta}_{(p)}\right)'\epsilon_0,$$

and

$$\begin{aligned} Var\left(\hat{\mathbf{y}}_0\right) &= Var\left(\hat{\mathbf{B}}'\mathbf{x}_0\right) + Var\left(\boldsymbol{\epsilon}_0\right) \\ &= \mathbf{x}_0'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{x}_0\Sigma + \Sigma \\ &= \left(\mathbf{x}_0'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{x}_0 + 1\right)\Sigma. \end{aligned}$$

By the fact that $E\left[\hat{\mathbf{y}}_0\right] = B'x_0$ and $\hat{\beta}_{(i)}'$s are multi- normally distributed, we know that

$$\hat{\mathbf{y}}_0 \sim N\left(\mathbf{B}'\mathbf{x}_0, \left(\mathbf{x}_0'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{x}_0 + 1\right)\Sigma\right).$$

(b) Note that by the same argument as above, we have

$$E\left[\mathbf{y}_0\right] = \widehat{\mathbf{B}}'\mathbf{x}_0 \sim N\left(\mathbf{B}'\mathbf{x}_0, \mathbf{x}_0'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{x}_0\Sigma\right),$$

and thus

$$\left(\boldsymbol{x}_0'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\boldsymbol{x}_0\right)^{-1/2}\Sigma^{-1/2}\left(E\left[\mathbf{y}_0\right] - \mathbf{B}'\mathbf{x}_0\right) \sim N\left(0, I_p\right).$$

Note that

$$\Sigma^{-1/2}E\Sigma^{-1/2} = \Sigma^{-1/2}\left(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}\right)'\left(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}\right)\Sigma^{-1/2} = \Sigma^{-1/2}\boldsymbol{\Xi}'\left(I - H\right)\boldsymbol{\Xi}\Sigma^{-1/2},$$

where $H = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$ and rank $(I - H) = n - q - 1$. By spectral decomposition, we have

$$I - H = Q \begin{pmatrix} I_{n-q-1} & 0 \\ 0 & 0 \end{pmatrix} Q',$$

where $Q = (q_{ij})_{n\times n}$ is orthogonal. So,

$$\Sigma^{-1/2}E\Sigma^{-1/2} = \left(\Sigma^{-1/2}\epsilon_1, \Sigma^{-1/2}\epsilon_2, \cdots, \Sigma^{-1/2}\epsilon_n\right) Q \begin{pmatrix} I_{n-q-1} & 0 \\ 0 & 0 \end{pmatrix} Q' \begin{pmatrix} \epsilon_1'\Sigma^{-1/2} \\ \epsilon_2'\Sigma^{-1/2} \\ \vdots \\ \epsilon_n'\Sigma^{-1/2} \end{pmatrix},$$

where ϵ_i is the ith row of Ξ and denoted by column vector. Now, $\Sigma^{-1/2}\epsilon_i \sim N(0, I_p)$ and $Cov\left(\Sigma^{-1/2}\boldsymbol{\epsilon}_i, \Sigma^{-1/2}\boldsymbol{\epsilon}_j\right) = 0$ for $i \neq j$. Now, denote

$$(e_1, e_2, \cdots, e_n) = \left(\Sigma^{-1/2}\epsilon_1, \Sigma^{-1/2}\epsilon_2, \cdots, \Sigma^{-1/2}\epsilon_n\right) Q,$$

we have

$$Cov(e_i, e_j) = Cov\left(\sum_{s=1}^{n} q_{si}\Sigma^{-1/2}\epsilon_s, \sum_{t=1}^{n} q_{tj}\Sigma^{-1/2}\epsilon_t\right) = \sum_{s=1}^{n} q_{si}q_{sj}Cov(\epsilon_s, \epsilon_s) = \delta_{ij}I_p,$$

where $\delta_{ij} = 1$ if $i = j; \delta_{ij} = 0$ otherwise. Now, we can conclude that e_i's are i.i.d. and $e_i \sim N(0, I_p)$. Thus,

$$\Sigma^{-1/2}E\Sigma^{-1/2} = (e_1, e_2, \cdots, e_n) \begin{pmatrix} I_{n-q-1} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} e_1' \\ e_2' \\ \vdots \\ e_n' \end{pmatrix} = \sum_{i=1}^{n-q-1} e_ie_i' \sim W_p(n - q - 1, I_p).$$

Therefore, according to the definition of Hotelling t distribution, we have

$$\left(E[\mathbf{y}_0] - \mathbf{B}'\mathbf{x}_0\right)' \Sigma^{-1/2} \left(\mathbf{x}_0'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{x}_0\right)^{-1/2} \left(\frac{\Sigma^{-1/2}E\Sigma^{-1/2}}{n-q-1}\right)^{-1} \left(\mathbf{x}_0'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{x}_0\right)^{-1/2} \Sigma^{-1/2}\left(E[\mathbf{y}_0] - \mathbf{B}'\mathbf{x}_0\right)$$
$$\sim T^2(n - q - 1),$$

i.e.,

$$\left(E[\mathbf{y}_0] - \mathbf{B}'\mathbf{x}_0\right)' \left(\frac{\mathbf{x}_0'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{x}_0 E}{n-q-1}\right)^{-1} \left(E[\mathbf{y}_0] - \mathbf{B}'\mathbf{x}_0\right) \sim T^2(n - q - 1).$$

Therefore, the $100(1 - \alpha)\%$ confidence ellipse is

$$\left(E[\mathrm{y}_0] - \mathrm{B}'\mathrm{x}_0\right)' \left(\frac{E}{n-q-1}\right)^{-1} \left(E[\mathrm{y}_0] - \mathrm{B}'\mathrm{x}_0\right) \leq T_\alpha^2(n - q - 1)\left(\mathrm{x}_0'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathrm{x}_0\right).$$

2 Obtain Principle Components

(a) To obtain principle components directly by inequalities, we denote $(\hat{\lambda}_1, \hat{e}_1), (\hat{\lambda}_2, \hat{e}_2), \ldots, (\hat{\lambda}_p, \hat{e}_p)$ to be the standardized ordered eigen-pairs of Σ. Because they formed a set of standard orthogonal bases on the $\mathbb{R}^p$ plane, for any vector $a_i \in \mathbb{R}^p$, it can be represented by the linear combination of eigenvectors, i.e. $a_i = \sum_{i=1}^p c_i \hat{e}_i$. While the constraint $a_i' a_i = 1$ is equivalent to $\sum_{i=1}^p c_i^2 = 1$. Then for the first principle component $a_1 y$:

$$
\begin{aligned}
Var(a_1 y) &= a_1' \Sigma a_1 \\
&= \left(\sum_{i=1}^p c_i \hat{e}_i \right)' \Sigma \left(\sum_{i=1}^p c_i \hat{e}_i \right) \\
&= \sum_{i=1}^p \hat{\lambda}_i c_i^2 \leq \hat{\lambda}_1.
\end{aligned}
$$

Under the constraint that $\sum_{i=1}^p c_i^2 = 1$. Hence $Var(a_1 y)$ is maximized when $a_1 = \hat{e}_1$.

For the j-th principle component $a_j y$, the constraints are equivalent to $c_1 = c_2 = \cdots = c_{j-1} = 0, \sum_{i=1}^p c_i^2 = 1$, and

$$
\begin{aligned}
Var(a_j y) &= a_j' \Sigma a_j \\
&= \left(\sum_{i=1}^p c_i \hat{e}_i \right)' \Sigma \left(\sum_{i=1}^p c_i \hat{e}_i \right) \\
&= \sum_{i=1}^p \hat{\lambda}_i c_i^2 \leq \hat{\lambda}_j.
\end{aligned}
$$

Therefore, we can obtain principle components simply by properties of the quadratic form of Σ.

(b) The Lagrange multiplier method can be used to incorporate the constraint $a_i' a_i = 1$ directly into the optimization problem.

Form the objective function of Lagrangian:

$$\mathcal{L}(\mathbf{a_i}, \lambda) = \mathbf{a_i}' \Sigma \mathbf{a_i} - \lambda(\mathbf{a_i}' \mathbf{a_i} - 1)$$

Then we take partial derivatives with respect to $\mathbf{a_i}$ and λ, and set them to zero:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{a_i}} = 2\Sigma \mathbf{a_i} - 2\lambda \mathbf{a_i} = 0 \implies \Sigma \mathbf{a_i} = \lambda \mathbf{a_i}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \mathbf{a_i}' \mathbf{a_i} - 1 = 0$$

This leads to the eigenvalue problem $\Sigma \mathbf{a_i} = \lambda \mathbf{a_i}$. By (a), the eigenvector corresponding to the largest eigenvalue λ is the principal component.

(c)

3 Calculating Principle Components

(a) For

$$\Sigma = \begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix},$$

the eigen-pairs of Σ are

$$\lambda_1 = 6, e_1 = (2\sqrt{5}/5, \sqrt{5}/5)',$$

$$\lambda_2 = 1, e_2 = (\sqrt{5}/5, -2\sqrt{5}/5)'.$$

Thus the population principle components $\mathbf{z} = (Z_1, Z_2)'$ are:

$$Z_1 = \frac{2\sqrt{5}}{5} Y_1 + \frac{\sqrt{5}}{5} Y_2,$$

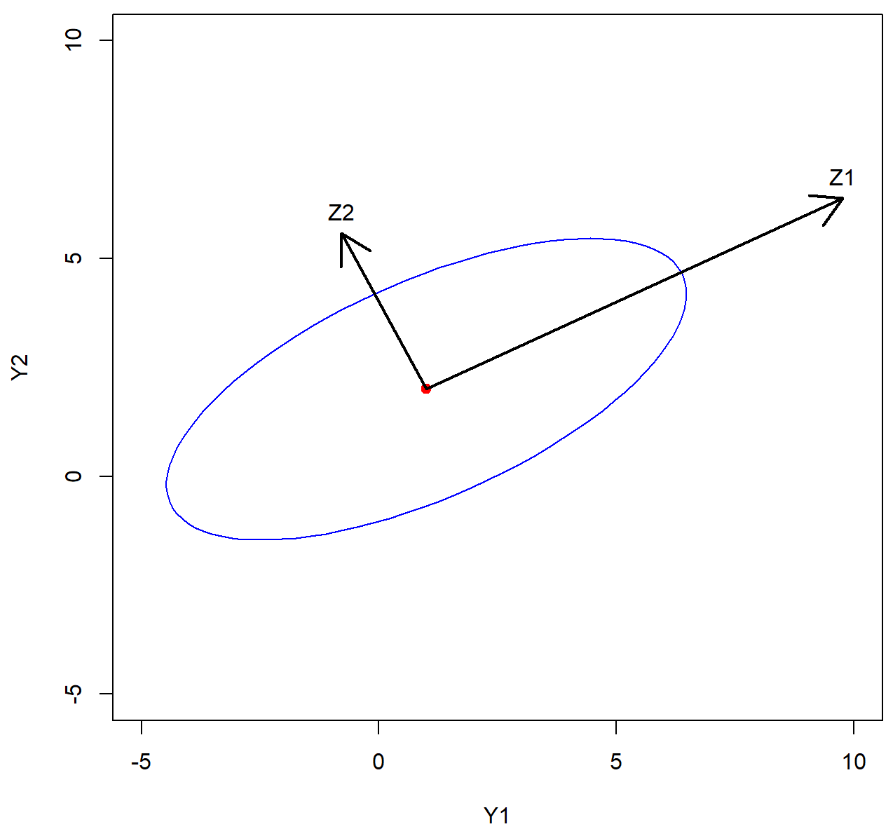
$$Z_2 = \frac{\sqrt{5}}{5} Y_1 - \frac{2\sqrt{5}}{5} Y_2.$$

(b) The proportion of total population variance explained by the first principal component is given by

$$\frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{6}{7}.$$

(c)

```r
library(ellipse)
library(plotrix)
mean_vector <- c(1,2)
cov_matrix <- matrix(c(5,2,2,2), nrow = 2)
ellipse_points <- ellipse(cov_matrix, centre = mean_vector, level = 0.95)
plot(NA, xlim = c(-5, 10), ylim = c(-5, 10), xlab = 'Y1', ylab = 'Y2')
lines(ellipse_points, type = 'l', col = 'blue')
points(mean_vector[1], mean_vector[2], col = 'red', pch = 19)
eigen_decomp <- eigen(cov_matrix)
eigenvalues <- eigen_decomp$values
eigenvectors <- - eigen_decomp$vectors
long_axis_length <- 4 * sqrt(eigenvalues[1])
short_axis_length <- 4 * sqrt(eigenvalues[2])
arrows(mean_vector[1], mean_vector[2], mean_vector[1] + eigenvectors[1,1] *
↪ long_axis_length, mean_vector[2] + eigenvectors[2,1] * long_axis_length, col =
↪ 'black', lwd = 2)
arrows(mean_vector[1], mean_vector[2], mean_vector[1] + eigenvectors[1,2] *
↪ short_axis_length, mean_vector[2] + eigenvectors[2,2] * short_axis_length, col
↪ = 'black', lwd = 2)
text(mean_vector[1] + eigenvectors[1,1] * long_axis_length, mean_vector[2] +
↪ eigenvectors[2,1] * long_axis_length, labels = "Z1", pos = 3, col = 'black')
text(mean_vector[1] + eigenvectors[1,2] * short_axis_length, mean_vector[2] +
↪ eigenvectors[2,2] * short_axis_length, labels = "Z2", pos = 3, col = 'black')
```



(d)

$$P = D_s^{-1}\Sigma D_s^{-1} = \begin{pmatrix} 1 & \sqrt{2}/\sqrt{5} \\ \sqrt{2}/\sqrt{5} & 1 \end{pmatrix} \implies |\tilde{\lambda}I - \Sigma| = \tilde{\lambda}^2 - 2\tilde{\lambda} + \frac{3}{5} \implies \tilde{\lambda}_1 = 1 + \frac{\sqrt{2}}{\sqrt{5}}, \tilde{\lambda}_2 = 1 - \frac{\sqrt{2}}{\sqrt{5}},$$

$$\begin{pmatrix} \tilde{\lambda}_1 - 1 & -\sqrt{2}/\sqrt{5} \\ -\sqrt{2}/\sqrt{5} & \tilde{\lambda}_1 - 1 \end{pmatrix} \tilde{e}_1 = 0 \quad \text{s.t.} \quad \tilde{e}_1'\tilde{e}_1 = 1 \implies \tilde{e}_1 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \implies V_1 = \frac{1}{\sqrt{2}}W_1 + \frac{1}{\sqrt{2}}W_2,$$

$$\begin{pmatrix} \tilde{\lambda}_2 - 1 & -\sqrt{2}/\sqrt{5} \\ -\sqrt{2}/\sqrt{5} & \tilde{\lambda}_2 - 1 \end{pmatrix} \tilde{e}_2 = 0 \quad \text{s.t.} \quad \tilde{e}_2'\tilde{e}_2 = 1 \implies \tilde{e}_2 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix} \implies V_2 = \frac{1}{\sqrt{2}}W_1 - \frac{1}{\sqrt{2}}W_2.$$

(e) When performing Principal Component Analysis (PCA) using the covariance matrix Σ and the correlation matrix $\mathbf{P}$, the proportion of variance explained by the first principal component is usually different. Here's why:

1. PCA based on the covariance matrix Σ uses the original scale of the data.
2. The principal components derived from Σ primarily reflect the variation in the data in its original units.
3. PCA based on the correlation matrix $\mathbf{P}$ uses standardized data.

4. When computing the correlation matrix, the data is first standardized (each variable is centered by subtracting its mean and then scaled by its standard deviation), so all variables have the same variance (i.e., 1).

5. The principal components derived from $\mathbf{P}$ primarily reflect the variation in the standardized data.

Since the covariance matrix and the correlation matrix are constructed differently, with the covariance matrix Σ reflecting the absolute variation in the data and the correlation matrix $\mathbf{P}$ reflecting the relative variation in the standardized data, the proportion of variance explained by the first principal component is generally not the same when using these two matrices for PCA.

Specifically, if the variables in the data have different scales and units, the covariance matrix Σ will be affected by these scales and units, whereas the correlation matrix $\mathbf{P}$ will not. Even if all variables have the same scale and units, the standardization process in computing the correlation matrix can lead to differences in the proportion of explained variance.

(f)

$$corr(z,y) = corr\left(\begin{pmatrix} e_1' \\ e_2' \end{pmatrix} y, y\right) = \begin{pmatrix} \lambda_1^{-1/2} & 0 \\ 0 & \lambda_2^{-1/2} \end{pmatrix} \begin{pmatrix} e_1' \\ e_2' \end{pmatrix} \sum_{j=1}^{2} \lambda_j e_j e_j' D_s^{-1} = \begin{pmatrix} \lambda_1^{1/2} e_1' \\ \lambda_2^{1/2} e_2' \end{pmatrix} D_s^{-1} = \begin{pmatrix} 2\sqrt{6}/5 & \sqrt{3}/\sqrt{5} \\ 1/5 & -2/\sqrt{5} \end{pmatrix},$$

$$\begin{aligned} corr(v,y) = corr\left(\begin{pmatrix} \tilde{e}_1' \\ \tilde{e}_2' \end{pmatrix} w, D_s w + \mu\right) &= \begin{pmatrix} \tilde{\lambda}_1^{-1/2} & 0 \\ 0 & \tilde{\lambda}_2^{-1/2} \end{pmatrix} \begin{pmatrix} \tilde{e}_1' \\ \tilde{e}_2' \end{pmatrix} \sum_{j=1}^{2} \tilde{\lambda}_j \tilde{e}_j \tilde{e}_j' D_s D_s^{-1} \\ &= \begin{pmatrix} \tilde{\lambda}_1^{1/2} \tilde{e}_1' \\ \tilde{\lambda}_2^{1/2} \tilde{e}_2' \end{pmatrix} = \begin{pmatrix} \tilde{\lambda}_1^{1/2}/\sqrt{2} & \tilde{\lambda}_1^{1/2}/\sqrt{2} \\ \tilde{\lambda}_2^{1/2}/\sqrt{2} & -\tilde{\lambda}_2^{1/2}/\sqrt{2} \end{pmatrix}, \end{aligned}$$

$$\begin{aligned} corr(z,v) &= corr\left(\begin{pmatrix} e_1' \\ e_2' \end{pmatrix} y, \begin{pmatrix} \tilde{e}_1' \\ \tilde{e}_2' \end{pmatrix} D_s^{-1}(y-\mu)\right) = \begin{pmatrix} \lambda_1^{-1/2} & 0 \\ 0 & \lambda_2^{-1/2} \end{pmatrix} \begin{pmatrix} e_1' \\ e_2' \end{pmatrix} \sum_{j=1}^{2} \lambda_j e_j e_j' D_s^{-1} (\tilde{e}_1, \tilde{e}_2) \begin{pmatrix} \tilde{\lambda}_1^{-1/2} & 0 \\ 0 & \tilde{\lambda}_2^{-1/2} \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1^{1/2} e_1' \\ \lambda_2^{1/2} e_2' \end{pmatrix} D_s^{-1} (\tilde{\lambda}_1^{1/2} \tilde{e}_1, \tilde{\lambda}_2^{1/2} \tilde{e}_2) \begin{pmatrix} \tilde{\lambda}_1^{-1} & 0 \\ 0 & \tilde{\lambda}_2^{-1} \end{pmatrix} = corr(z,y) corr(y,v) \begin{pmatrix} \tilde{\lambda}_1^{-1} & 0 \\ 0 & \tilde{\lambda}_2^{-1} \end{pmatrix}. \end{aligned}$$

Thus, there might exist sort of transitivity property among z, y and v.

4 Factor Analysis

(a) To obtain the estimated factor loadings by principle components method with three factors,

```r
rm(list = ls())
R = matrix(c(1,0.505,0.569,0.602,0.621,0.603,0.505,1,0.422,
             0.467,0.482,0.450,0.569,0.422,1,0.926,0.877,0.878,
             0.602,0.467,0.926,1,0.874,0.894,0.621,0.482,0.877,
             0.874,1,0.937,0.603,0.450,0.878,0.894,0.937,1),
           nrow = 6,ncol = 6,byrow = TRUE)
eigenvalues <- eigen(R)$values
```

```r
eigenvectors <- eigen(R)$vectors
L <- eigenvectors[, 1:3] %*% diag(sqrt(eigenvalues[1:3]))
L
```

```
           [,1]       [,2]        [,3]
[1,] -0.7405352  0.3499708  0.57325558
[2,] -0.6041852  0.7205817 -0.33990980
[3,] -0.9288076 -0.2328502 -0.07482720
[4,] -0.9433880 -0.1744338 -0.06700492
[5,] -0.9476006 -0.1428236 -0.04459705
[6,] -0.9446719 -0.1888052 -0.04676285
```

(b) The estimated specific variances are:

```r
Psi <- diag(diag(R - L %*% t(L)))
Psi
```

```
             [,1]         [,2]       [,3]       [,4]       [,5]       [,6]
[1,] 0.0005060766 0.0000000000 0.00000000 0.00000000 0.00000000 0.00000000
[2,] 0.0000000000 0.0001835913 0.00000000 0.00000000 0.00000000 0.00000000
[3,] 0.0000000000 0.0000000000 0.07749811 0.00000000 0.00000000 0.00000000
[4,] 0.0000000000 0.0000000000 0.00000000 0.07510225 0.00000000 0.00000000
[5,] 0.0000000000 0.0000000000 0.00000000 0.00000000 0.07966566 0.00000000
[6,] 0.0000000000 0.0000000000 0.00000000 0.00000000 0.00000000 0.06976083
```

(c) The estimated communalities are:

```r
H <- diag(L%*%t(L))
H
```

```
[1] 0.9994939 0.9998164 0.9225019 0.9248977 0.9203343 0.9302392
```

(d)

```r
H / sum(eigenvalues)
```

```
[1] 0.1665823 0.1666361 0.1537503 0.1541496 0.1533891 0.1550399
```

(e) The residual matrix is:

```r
print(round(R-L%*%t(L)-psi, 5))
```

```
        [,1]     [,2]     [,3]     [,4]     [,5]     [,6]
[1,]  0.00000  0.00025  0.00557  0.00285 -0.00518 -0.00368
[2,]  0.00025  0.00000  0.00318 -0.00006 -0.00277 -0.00060
[3,]  0.00557  0.00318  0.00000  0.00414 -0.03973 -0.04688
[4,]  0.00285 -0.00006  0.00414  0.00000 -0.04786 -0.03326
[5,] -0.00518 -0.00277 -0.03973 -0.04786  0.00000  0.01278
[6,] -0.00368 -0.00060 -0.04688 -0.03326  0.01278  0.00000
```

Upon examining the residual matrix, we observe that all the entries are relatively small. This indicates that the variables can be adequately explained by three factors.

(f) F_1^*: age;F_2^*: skull breadth;F_3^*: skull length.

5 Mix-do Problem

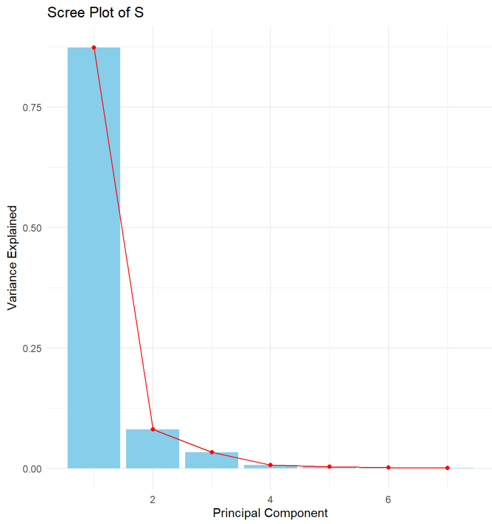
(a)

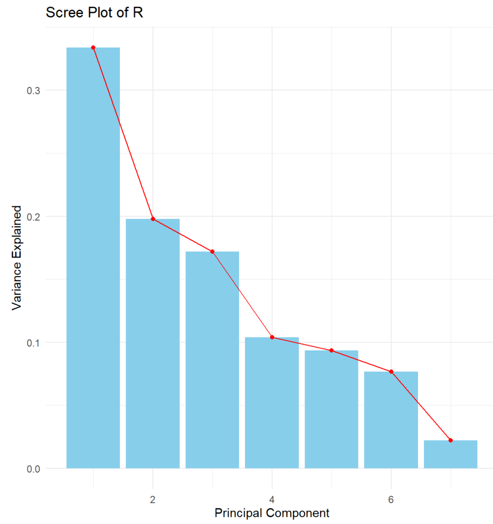
```r
rm(list = ls())
data <- read.table("C:/Users/Ray Chen/Desktop/pollution.dat")
S <- cov(data)
R <- cor(data)
S_eigenvalues <- round(eigen(S)$values,4)
R_eigenvalues <- round(eigen(R)$values,4)
S_eigenvalues
R_eigenvalues
```

```
> S_eigenvalues
[1] 304.2579  28.2761  11.4645   2.5243   1.2795   0.5287   0.2096
> R_eigenvalues
[1] 2.3368 1.3860 1.2041 0.7271 0.6535 0.5367 0.1559
```

```r
library(ggplot2)
library(cowplot)
S_explained_variance_proportion <- S_eigenvalues / sum(S_eigenvalues)
S_scree_data <- data.frame(
  Principal_Component = 1:length(S_explained_variance_proportion),
  Variance_Explained = S_explained_variance_proportion
```

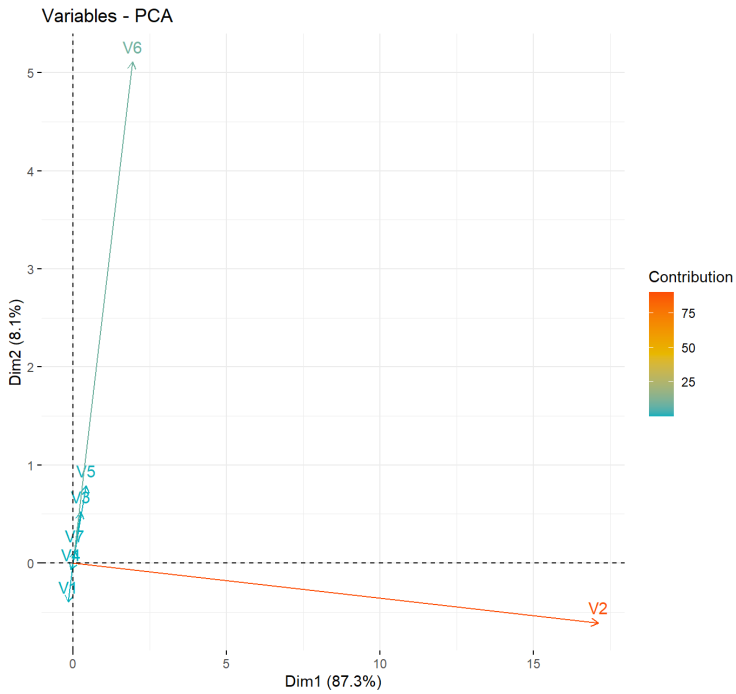
```
)
R_explained_variance_proportion <- R_eigenvalues / sum(R_eigenvalues)
R_scree_data <- data.frame(
  Principal_Component = 1:length(R_explained_variance_proportion),
  Variance_Explained = R_explained_variance_proportion
)
scree_plot_S <- ggplot(S_scree_data, aes(x = Principal_Component, y =
↪  Variance_Explained)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  geom_line(aes(x = Principal_Component, y = Variance_Explained), color = "red") +
  geom_point(aes(x = Principal_Component, y = Variance_Explained), color = "red")
  ↪  +
  labs(title = "Scree Plot of S",
       x = "Principal Component",
       y = "Variance Explained") +
  theme_minimal()
scree_plot_R <- ggplot(R_scree_data, aes(x = Principal_Component, y =
↪  Variance_Explained)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  geom_line(aes(x = Principal_Component, y = Variance_Explained), color = "red") +
  geom_point(aes(x = Principal_Component, y = Variance_Explained), color = "red")
  ↪  +
  labs(title = "Scree Plot of R",
       x = "Principal Component",
       y = "Variance Explained") +
  theme_minimal()
combined_plot <- plot_grid(scree_plot_S, scree_plot_R, ncol = 2)
print(combined_plot)
```

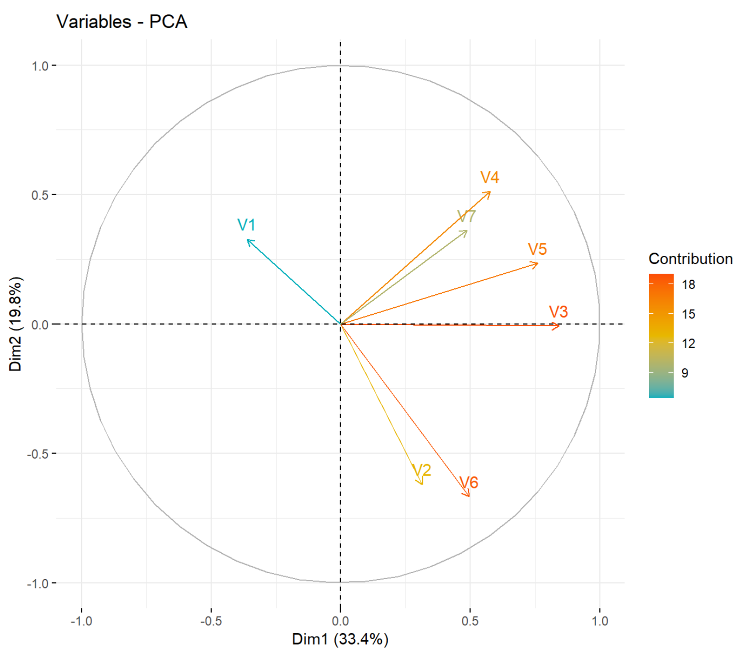




Checking the scree plot of both covariance matrix and correlation matrix, we choose two factor for covariance-based method and six for correlation-based method.

```r
library(FactoMineR)
library(factoextra)
pca_result_S <- PCA(data, graph = FALSE,ncp = 2, scale.unit = FALSE)
fviz_pca_var(pca_result_S,
             col.var = "contrib",
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
             legend.title = "Contribution")
pca_result_R <- PCA(data, graph = FALSE,ncp = 6, scale.unit = TRUE)
fviz_pca_var(pca_result_R,
             col.var = "contrib",
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
             legend.title = "Contribution")
```





When using covariance matrix:

The first principal component (Dim1) explains 87.3% of the variance. The second principal component (Dim2) explains 8.1% of the variance. Together, these two principal components explain 95.4% of the variance.

Variable V_2 has the largest positive contribution to the first principal component. Variable V_6 has the largest positive contribution to the second principal component. Other variables V_1, V_3, V_4, V_5 are clustered near the origin, indicating they have smaller contributions to these two principal components.

Variables V_2 and V_6 show significant independence in these two principal components. Other variables have smaller contributions and are clustered near the origin, indicating lower differentiation in these two principal components.

When using Correlation Matrix:

The first principal component (Dim1) explains 33.4% of the variance. The second principal component (Dim2) explains 19.8% of the variance. Together, these two principal components explain 53.2% of the variance.

All variables are relatively evenly distributed around the circumference, indicating a balanced contribution to the principal components. Variables V_4, V_5, V_3 contribute more to the first principal component. Variables V_2, V_6 contribute more to the second principal component.

All variables have a significant explanatory power in the first and second principal components, indicating a strong linear correlation between variables. The angle between variables represents their correlation; smaller angles indicate positive correlation, larger angles indicate negative correlation.

When using covariance matrix S, we focus on the original scale of variables, thus variables V_2 and V_6 have significant contributions due to their large values. Variables with larger variance have a greater influence on the principal components.

When using correlation matrix R, Variables are standardized, focusing on the relative relationships between variables rather than their absolute values. Variable V_1 has the smallest contribution to the first principal component, indicating weaker explanatory power in the correlation matrix.

In summary, using the covariance matrix results in variance being concentrated in a few variables, whereas using the correlation matrix results in more balanced contributions from all variables. The choice of matrix depends on the specific analysis needs and characteristics of the data.

(b)

```r
library(psych)
y <- data[c(1,2,5,6)]
principal(y, nfactors=1, rotate="none", covar=F)
principal(y, nfactors=2, rotate="none", covar=F)
```

```
> principal(y, nfactors=1, rotate="none", covar=F)
Principal Components Analysis
Call: principal(r = y, nfactors = 1, rotate = "none", covar = F)
Standardized loadings (pattern matrix) based upon correlation matrix
     PC1   h2   u2 com
V1 -0.56 0.32 0.68   1
V2  0.65 0.42 0.58   1
V5  0.48 0.23 0.77   1
```

```
V6  0.77 0.59 0.41   1

                PC1
SS loadings    1.56
Proportion Var 0.39

Mean item complexity =  1
Test of the hypothesis that 1 component is sufficient.

The root mean square of the residuals (RMSR) is  0.2
 with the empirical chi square  19.81  with prob <  5e-05

Fit based upon off diagonal values = -0.03
> principal(y, nfactors=2, rotate="none", covar=F)
Principal Components Analysis
Call: principal(r = y, nfactors = 2, rotate = "none", covar = F)
Standardized loadings (pattern matrix) based upon correlation matrix
     PC1   PC2   h2   u2 com
V1 -0.56  0.24 0.38 0.62 1.4
V2  0.65  0.52 0.69 0.31 1.9
V5  0.48 -0.74 0.77 0.23 1.7
V6  0.77  0.20 0.63 0.37 1.1

                       PC1  PC2
SS loadings           1.56 0.91
Proportion Var        0.39 0.23
Cumulative Var        0.39 0.62
Proportion Explained  0.63 0.37
Cumulative Proportion 0.63 1.00

Mean item complexity =  1.5
Test of the hypothesis that 2 components are sufficient.

The root mean square of the residuals (RMSR) is  0.21
 with the empirical chi square  22.56  with prob <  NA

Fit based upon off diagonal values = -0.17
```

(c)

```r
library(GPArotation)
fa(y, nfactors = 1, fm = "ml")
fa(y, nfactors = 2, fm = "ml")
```

```
> fa(y, nfactors = 1, fm = "ml")
Factor Analysis using method =  ml
Call: fa(r = y, nfactors = 1, fm = "ml")
Standardized loadings (pattern matrix) based upon correlation matrix
     ML1    h2   u2 com
V1 -0.32 0.105 0.89   1
V2  0.41 0.168 0.83   1
V5  0.23 0.054 0.95   1
V6  0.77 0.595 0.41   1

                ML1
SS loadings    0.92
Proportion Var 0.23

Mean item complexity =  1
Test of the hypothesis that 1 factor is sufficient.

df null model =  6  with the objective function =  0.21 with Chi Square =  8.23
df of  the model are 2  and the objective function was  0

The root mean square of the residuals (RMSR) is  0.02
The df corrected root mean square of the residuals is  0.04

The harmonic n.obs is  42 with the empirical chi square  0.23  with prob <  0.89
The total n.obs was  42  with Likelihood Chi Square =  0.15  with prob <  0.93

Tucker Lewis Index of factoring reliability =  3.666
RMSEA index =  0  and the 90 % confidence intervals are  0 0.093
BIC =  -7.33
Fit based upon off diagonal values = 0.99
Measures of factor score adequacy
                                                   ML1
Correlation of (regression) scores with factors   0.81
Multiple R square of scores with factors          0.65
Minimum correlation of possible factor scores     0.30
```

```
> fa(y, nfactors = 2, fm = "ml")
Factor Analysis using method =  ml
Call: fa(r = y, nfactors = 2, fm = "ml")
Standardized loadings (pattern matrix) based upon correlation matrix
    ML1   ML2    h2   u2 com
V1 0.01  0.50 0.248 0.75 1.0
V2 0.60  0.08 0.323 0.68 1.0
V5 0.15 -0.15 0.067 0.93 2.0
V6 0.45 -0.30 0.425 0.58 1.7

                       ML1  ML2
SS loadings           0.64 0.42
Proportion Var        0.16 0.11
Cumulative Var        0.16 0.27
Proportion Explained  0.60 0.40
Cumulative Proportion 0.60 1.00

 With factor correlations of
      ML1   ML2
ML1  1.00 -0.49
ML2 -0.49  1.00

Mean item complexity =  1.4
Test of the hypothesis that 2 factors are sufficient.

df null model =  6  with the objective function =  0.21 with Chi Square =  8.23
df of  the model are -1  and the objective function was  0

The root mean square of the residuals (RMSR) is  0
The df corrected root mean square of the residuals is  NA

The harmonic n.obs is  42 with the empirical chi square  0  with prob <  NA
The total n.obs was  42  with Likelihood Chi Square =  0  with prob <  NA

Tucker Lewis Index of factoring reliability =  4.084
Fit based upon off diagonal values = 1
Measures of factor score adequacy
                                                   ML1   ML2
Correlation of (regression) scores with factors   0.73  0.65
```

```
Multiple R square of scores with factors          0.53  0.43
Minimum correlation of possible factor scores     0.05 -0.15
```

(d) Based on principal component method, the loadings are:

$$L = (-.56, .65, .48, .77)'.$$

Based on MLE, the loadings are:

$$L = (-.32, .41, .23, .77)'.$$

So, based on different methods, we can obtain different results.

(e)

```r
fac <- principal(y, nfactors=2, rotate="varimax", covar=F)
fac$loadings
```

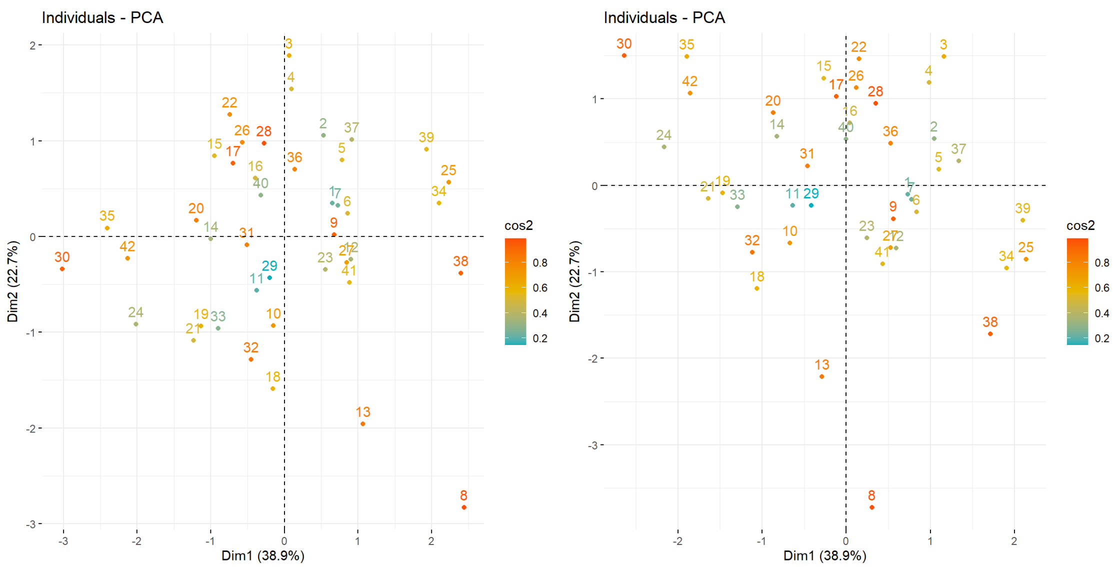
```
Loadings:
   RC1    RC2
V1 -0.313 -0.528
V2  0.828
V5         0.875
V6  0.739  0.295

                 RC1   RC2
SS loadings    1.332 1.133
Proportion Var 0.333 0.283
Cumulative Var 0.333 0.616
```

We can see that RC1 and RC2 can account for 61.6% of the variance. And RC1 can be explained as the strength of the sun, and RC2 can be explained as the intensity of NO2.

(f)

```r
fa_result <- PCA(y, ncp = 2, graph = FALSE)
loadings <- fa_result$var$coord
rotation <- varimax(loadings)
rotated_loadings <- rotation$loadings
original_scores <- fa_result$ind$coord
rotation_matrix <- rotation$rotmat
rotated_scores <- as.matrix(original_scores) %*% rotation_matrix
rotated_fa_result <- fa_result
rotated_fa_result$ind$coord <- rotated_scores
```

```
p1 <- fviz_pca_ind(fa_result, col.ind = "cos2", gradient.cols = c("#00AFBB",
↪  "#E7B800", "#FC4E07"))
p2 <- fviz_pca_ind(rotated_fa_result, col.ind = "cos2", gradient.cols =
↪  c("#00AFBB", "#E7B800", "#FC4E07"))
pp <- plot_grid(p1,p2, ncol = 2)
print(pp)
```



The points in the first (third) quadrant face the strong(weak) sun and intensive(thin) NO2; the points in the second (forth) quadrant face the weak(strong) sun and intensive(thin) NO2.