Multivariate Statistics

Chapter 3: Principal Component Analysis

Pedro Galeano
Departamento de Estadística
Universidad Carlos III de Madrid
pedro.galeano@uc3m.es

Course 2017/2018

Master in Mathematical Engineering

Introduction

- Suppose that we have a data matrix X with dimension $n \times p$.
- A central problem in multivariate data analysis is the curse of dimensionality: if the ratio n/p is not large enough, some problems might be intractable.
- For example, assume that we have a sample of size n from a p-dimensional random variable following a $N(\mu_x, \Sigma_x)$ distribution.
- In this case, the number of parameters to estimate is $p + p(p+1)/2$.
- For instance, for $p = 5$ and $p = 10$, there are 20 and 65 parameters, respectively.
- Thus, the larger p, the larger number of observations we need to obtain reliable estimates of the parameters.

Introduction

- There are several dimension reduction techniques that try to answer the same question:
  - Is it possible to describe with accuracy the values of p variables with a smaller number $r < p$ of new variables?
- We are going to see in this chapter principal component analysis.
- Next chapter is devoted to factor analysis.

Introduction

- As mentioned before, the main objective of principal component analysis (PCA) is to reduce the dimension of the problem.
- The simplest way of dimension reduction is to take just some variables of the observed multivariate random variable $x = (x_1, \ldots, x_p)'$ and to discard all others.
- However, this is not a very reasonable approach since we loss all the information contained in the discarded variables.
- Principal component analysis is a flexible approach based on a few linear combinations of the original (centered) variables in $x = (x_1, \ldots, x_p)'$.
- The p components of x are required to reproduce the total system variability.
- However much of the variability of x can be accounted for a small number of $r < p$ of principal components.
- If so, there is almost as much information in the r principal components as there is in the original p variables contained in x.

Introduction

- The general objective of PCA is dimension reduction.
- However, PCA is a powerful method to interpret the relationship between the univariate variables that form $x = (x_1, \ldots, x_p)'$.
- Indeed, a PCA often reveals relationships that were not previously suspected and thereby allows interpretations that would not ordinarily results.
- It is important to note that a PCA is more of a means to an end rather than an end in themselves, because they frequently serve as intermediate steps in much larger investigations.
- As we shall see, principal components depend solely on the covariance (or correlation) matrix of x.
- Therefore, their development does not require a multivariate Gaussian assumption.

Principal component analysis

- Given a multivariate random variable $x = (x_1, \ldots, x_p)'$ with mean μ_x and covariance matrix Σ_x, the principal components are contained in a new multivariate random variable of dimension $r \leq p$ given by:

$$z = A' (x - \mu_x)$$

  where A is a certain $p \times r$ matrix whose columns are $p \times 1$ vectors $a_j = (a_{j1}, \ldots, a_{jp})'$, for $j = 1, \ldots, r$.

- Therefore, $z = (z_1, \ldots, z_r)'$ is a linear transformation of x given by the r linear combinations a_j, for $j = 1, \ldots, r$.

Principal component analysis

- Consequently:

$$E(z_j) = 0$$
$$Var(z_j) = a_j' \Sigma_x a_j$$
$$Cov(z_j, z_k) = a_j' \Sigma_x a_k$$

for $j, k = 1, \ldots, r$.

Principal component analysis

- In particular, among all the possible linear combinations of the variables in $x = (x_1, \ldots, x_p)'$, the principal components are those that simultaneously verifies the following two properties:

  1. The variances $Var(z_j) = a_j' \Sigma_x a_j$ are as large as possible.
  2. The covariances $Cov(z_j, z_k) = a_j' \Sigma_x a_k$ are 0, implying that the principal components of x, i.e., the variables in $z = (z_1, \ldots, z_r)'$, are uncorrelated.

- Next, we formally derive the linear combinations a_j, for $j = 1, \ldots, r$ that leads to the PC's.

Principal component analysis

- Assume that the covariance matrix Σ_x have a set of p eigenvalues $\lambda_1, \ldots, \lambda_p$ with associated eigenvectors $v_1, \ldots, v_p$.
- Then, the first principal component corresponds to the linear combination with maximum variance.
- In other words, the first PC corresponds to the linear combination that maximizes $Var(z_1) = \sigma_{z_1}^2 = a_1' \Sigma_x a_1$.
- However, it is clear that $a_1' \Sigma_x a_1$ can be increased by multiplying any a_1 with some constant.
- To eliminate this indeterminacy, it is convenient to restrict attention to coefficient vector of unit length, i.e., we assume that $a_1' a_1 = 1$.
- The following linear combinations are obtained with a similar argument but adding the property that they are uncorrelated with the previous ones.

Principal component analysis

- Therefore, we define:

$$\text{First principal component} = \underset{s.t.\ a_1' a_1 = 1}{\arg\max}\ a_1' \Sigma_x a_1$$

$$\text{Second principal component} = \underset{s.t.\ a_2' a_2 = 1,\ a_1' \Sigma_x a_2 = 0}{\arg\max}\ a_2' \Sigma_x a_2$$

$$\vdots$$

$$\text{r-th principal component} = \underset{s.t.\ a_r' a_r = 1,\ a_r' \Sigma_x a_k = 0,\ \text{for } k < r}{\arg\max}\ a_r' \Sigma_x a_r$$

Principal component analysis

- The first problem can be solved with the Lagrange multiplier method as follows.
- Let:

$$M = a_1' S_x a_1 - \beta_1 \left(a_1' a_1 - 1 \right)$$

- Then,

$$\frac{\partial M}{\partial a_1} = 2\Sigma_x a_1 - 2\beta_1 a_1 = 0 \iff \Sigma_x a_1 = \beta_1 a_1$$

- Therefore, a_1 is an eigenvector of Σ_x and β_1 is its corresponding eigenvalue.
- Which ones? Multiplying by a_1' in the last expression, we get:

$$\sigma_{z_1}^2 = a_1' \Sigma_x a_1 = \beta_1 a_1' a_1 = \beta_1$$

- As $\sigma_{z_1}^2$ should be maximal, β_1 corresponds to the largest eigenvalue of Σ_x ($\beta_1 = \lambda_1$) and a_1 is its associated eigenvector ($a_1 = v_1$).

Principal component analysis

- The second principal component is given by:

$$\text{Second principal component} = \underset{s.t.\ a_2' a_2 = 1,\ a_1' \Sigma_x a_2 = 0}{\arg\max} a_2' \Sigma_x a_2$$

  where $a_1 = v_1$, the eigenvector associated with the largest eigenvalue of the covariance matrix Σ_x.

- Therefore, $\Sigma_x v_1 = \lambda_1 v_1$, so that:

$$a_1' \Sigma_x a_2 = \lambda_1 v_1' a_2 = 0$$

- Following the reasoning for the first principal component, we define:

$$M = a_2' \Sigma_x a_2 - \beta_2 \left(a_2' a_2 - 1 \right)$$

- Then,

$$\frac{\partial M}{\partial a_2} = 2 \Sigma_x a_2 - 2 \beta_2 a_2 = 0 \iff \Sigma_x a_2 = \beta_2 a_2$$

Principal component analysis

- Therefore, a_2 is an eigenvector of Σ_x and β_2 is its corresponding eigenvalue.
- Which ones? Multiplying by a_2' in the last expression,

$$\sigma_{z_2}^2 = a_2' \Sigma_x a_2 = \beta_2 a_2' a_2 = \beta_2$$

- As $\sigma_{z_2}^2$ should be maximal, β_2 corresponds to the second largest eigenvalue of Σ_x ($\beta_2 = \lambda_2$) and a_2 is its associated eigenvector ($a_2 = v_2$).
- This argument can be extended for successive principal components.
- Therefore, the r principal components corresponds to the eigenvectors of the covariance matrix Σ_x associated with the r largest eigenvalues.

Principal component analysis

- In summary, the principal components are given by:

$$z = V_r' (x - \mu_x)$$

  where V_r is a $p \times r$ orthogonal matrix (i.e., $V_r' V_r = I_r$ and $V_r V_r' = I_p$) whose columns are the first r eigenvectors of Σ_x.

- The covariance matrix of z, Σ_z, is the diagonal matrix with elements $\lambda_1, \ldots, \lambda_r$, i.e., the first r eigenvalues of Σ_x.

- Therefore, the usefulness of the principal components is two-fold:
  - It allows for an optimal representation, in a space of reduced dimensions, of the original observations.
  - It allows the original correlated variables to be transformed into new uncorrelated variables, facilitating the interpretation of the data.

Principal component analysis

- Note that we can compute the principal components even for $r = p$.
- Taking this into account, it is easy to check that the PC's verifies the following properties:

$$
\begin{gathered}
E\left(z_j\right) = 0 \\
Var\left(z_j\right) = \lambda_j \\
Cov\left(z_j, z_k\right) = 0 \\
Var\left(z_1\right) \geq Var\left(z_2\right) \geq \cdots \geq Var\left(z_p\right) \geq 0 \\
\sum_{j=1}^{p} Var\left(z_j\right) = Tr\left(\Sigma_x\right) = \sum_{j=1}^{p} Var\left(x_j\right) \\
\prod_{j=1}^{p} Var\left(z_j\right) = \left|\Sigma_x\right|
\end{gathered}
$$

for $j, k = 1, \ldots, p$.

Principal component analysis

- In particular, note that the fourth property ensures that the set of p principal components conserve the initial variability.

- Therefore, a measure of how well the r-th PC explains variation is given by the proportion of variability explained by r-th PC is given by:

$$PV_r = \frac{\lambda_r}{\lambda_1 + \cdots + \lambda_p} \qquad r = 1, \ldots, p$$

- Additionally, a measure of how well the first r PCs explain variation is given by the accumulated proportion of variability explained by the first r PCs is given by:

$$APV_r = \frac{\lambda_1 + \cdots + \lambda_r}{\lambda_1 + \cdots + \lambda_p} \qquad r = 1, \ldots, p$$

- Therefore, if most (for instance, 80% or 90%) of the total variability can be attributed to the first one, two or three principal components, then these components can "replace" the original p variables without much loss of information.

Principal component analysis

- The covariance between the principal components and the original variables can be written as follows:

$$Cov\left(z,x\right) = E\left[z\left(x-\mu_x\right)'\right] = E\left[V_r'\left(x-\mu_x\right)\left(x-\mu_x\right)'\right] =$$
$$= V_r' E\left[\left(x-\mu_x\right)\left(x-\mu_x\right)'\right] = V_r'\Sigma_x$$

- Now, the singular value decomposition of Σ_x is given by $\Sigma_x = V_p \Lambda_p V_p'$, where Λ_p is the diagonal matrix that contains the p eigenvalues of Σ_x in decreasing order and V_p is the matrix that contains the p associated eigenvectors of Σ_x.

- Therefore:

$$Cov\left(z,x\right) = V_r' V_p \Lambda_p V_p' = \Lambda_r V_r'$$

  where Λ_r is the diagonal matrix that contains the r largest eigenvalues of Σ_x in decreasing order.

- In particular, note that $\Lambda_r = \Sigma_z$.

Principal component analysis

- On the other hand, the correlation between between the principal components and the original variables can be written as follows:

$$Cor\left(z,x\right)=\Sigma_z^{-1/2}E\left[z\left(x-\mu_x\right)'\right]D_x^{-1/2}=\Sigma_z^{-1/2}V_r'\Sigma_x D_x^{-1/2}$$

  where D_x is a $p\times p$ diagonal matrix whose elements are the variances in the main diagonal of Σ_x.

- Now, replacing Σ_x with $V_p\Lambda_p V_p'$ and Σ_z with Λ_r:

$$Cor\left(z,x\right)=\Lambda_r^{-1/2}V_r'V_p\Lambda_p V_p' D_x^{-1/2}=\Lambda_r^{1/2}V_r' D_x^{-1/2}$$

  because $\Lambda_r^{-1/2}V_r'V_p\Lambda_p V_p'=\Lambda_r^{1/2}V_r'$.

- The correlations of the principal components and the variables often help to interpret the components as they measure the contribution of each individual variable to each principal component.

Normalized principal component analysis

- One problem with principal components is that they are not scale-invariant because if we change the units of the variables, the covariance matrix of the transformed variables will also change.
- Additionally, if there are large differences between the variances of the original variables, then those whose variances are largest will tend to dominate the early components.
- In these circumstances, it is better first to standardize the variables.
- In other words, the principal components should only be extracted from the original variables when all of them have the same scale.

Normalized principal component analysis

- Therefore, if the variables have different units of measurement, we define $y = D_x^{-1/2}(x - \mu_x)$ as the univariate standardized original variables and obtain the PCs from y.

- For that, note that:

$$E(y) = 0_r$$

and

$$Cov(y) = E[yy'] = E\left[D_x^{-1/2}(x - \mu_x)(x - \mu_x)' D_x^{-1/2}\right] =$$
$$= D_x^{-1/2}\Sigma_x D_x^{-1/2} = \varrho_x$$

i.e., the covariance matrix of y is the correlation matrix of x.

Normalized principal component analysis

- Consequently, the principal components of y should be obtained from the eigenvectors of the correlation matrix of x, denoted by $v_1^\varrho, \ldots, v_r^\varrho$, with associated eigenvalues $\lambda_1^\varrho \geq \cdots \geq \lambda_r^\varrho \geq 0$.

- In particular, these are given by:

$$z^\varrho = (V_r^\varrho)' \, y = (V_r^\varrho)' \, D_x^{-1/2} (x - \mu_x)$$

  where $V_r^\varrho = [v_1^\varrho | \cdots | v_r^\varrho]$ is the $r \times r$ matrix that contains the eigenvectors of the correlation matrix ϱ_x.

- The PCs obtained in this way are called the normalized principal components.

Normalized principal component analysis

- All the previous results apply, with some simplifications as the variance of the univariate standardized variables is 1.
- Therefore, the normalized PC's verifies the following properties:

$$E\left(z_j^\varrho\right) = 0$$

$$Var\left(z_j^\varrho\right) = \lambda_j^\varrho$$

$$Cov\left(z_j^\varrho, z_k^\varrho\right) = 0$$

$$Var\left(z_1^\varrho\right) \geq Var\left(z_2^\varrho\right) \geq \cdots \geq Var\left(z_p^\varrho\right) \geq 0$$

$$\sum_{j=1}^{p} Var\left(z_j^\varrho\right) = Tr\left(\Sigma_y\right) = Tr\left(\varrho_x\right) = r$$

$$\prod_{j=1}^{p} Var\left(z_j^\varrho\right) = \left|\varrho_x\right|$$

for $j, k = 1, \ldots, p$.

Normalized principal component analysis

- For normalized PCs, the proportion of variability explained by r-th principal component is given by:

$$PV_r^{\varrho} = \frac{\lambda_r^{\varrho}}{p}$$

- Similarly, for normalized PCs, the accumulated proportion of variability explained by the first r principal components based is given by:

$$PV_r^{\varrho} = \frac{\lambda_1^{\varrho} + \cdots + \lambda_r^{\varrho}}{p}$$

Normalized principal component analysis

- Additionally, it is possible to show that the covariance between the principal components based on the standardized variables and the original variables can be written as follows:

$$Cov(z^{\varrho}, x) = (V_r^{\varrho})' D_x^{-1/2} \Sigma_x$$

- Moreover, the correlation between the principal components based on the standardized variables and the original variables can be written as follows:

$$Cor(z^{\varrho}, y) = (\Lambda_r^{\varrho})^{1/2} (V_r^{\varrho})'$$

- Here, Λ_r^{ϱ} is the covariance matrix of z, denoted by Σ_z^{ϱ}, that is a $r \times r$ diagonal matrix that contains the eigenvalues of ϱ_x, $\lambda_1^{\varrho}, \ldots, \lambda_r^{\varrho}$.

Principal component analysis in practice

- In practice, one replace the population quantities with their corresponding sample counterparts based on the data matrix X of dimension $n \times p$.
- The principal component scores are the values of the new variables.
- If we use the sample covariance matrix to obtain the principal components, the data matrix that contains the principal component scores is given by:

$$Z = \widetilde{X} V_r^{S_x}$$

where $V_r^{S_x}$ is the matrix that contains the eigenvectors of the sample covariance matrix S_x linked with the r largest eigenvalues.

Principal component analysis in practice

- On the other hand, if we use the sample correlation matrix to obtain the principal components, the data matrix that contains the principal component scores is given by:

$$Z = YV_r^{R_x} = \widetilde{X} D_{S_x}^{-1/2} V_r^{R_x}$$

where V_r^R is the matrix that contains the eigenvectors of the sample correlation matrix R_x linked with the r largest eigenvalues and D_{S_x} is the diagonal matrix that contains the sample variances of the components of X.

Principal component analysis in practice

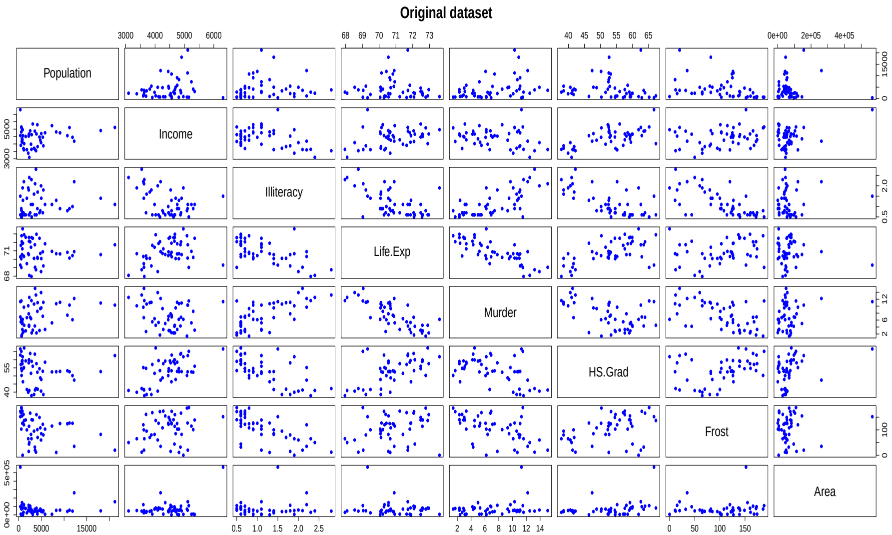
- Different rules have been suggested for selecting the number of components r for data sets:

  1. Plot a graph of $1, \ldots, p$ against $\lambda_1, \ldots, \lambda_p$ (the scree plot): The idea is to exclude of the analysis those components associated with small values and that approximately the same size.
  2. Select components until a certain proportion of the variance has been covered, such as 80% or 90%: This rule should be applied with caution because sometimes a single component picks up most of the variability, whereas there might be other components with interesting interpretations.
  3. Discard those components associated with eigenvalues of less than a certain value such as the mean of the eigenvalues of the sample covariance or correlation matrix: Again, this rule is arbitrary: a variable that is independent from the rest usually accounts for a principal component and can have a large eigenvalue.
  4. Use asymptotic results on the estimated eigenvalues of the covariance or correlation matrices used to derive the PCs.

Illustrative example I

- Consider the following eight univariate variables measured on the 50 states of the USA:
  - x_1: population estimate as of July 1, 1975 (in thousands).
  - x_2: per capita income (1974) (in dollars).
  - x_3: illiteracy (1970, percent of population).
  - x_4: life expectancy in years $(1969-71)$.
  - x_5: murder and non-negligent manslaughter rate per 100000 population (1976).
  - x_6: percent high-school graduates (1970).
  - x_7: mean number of days with minimum temperature below freezing (1931-1960) in capital or large city.
  - x_8: land area in square miles.

Illustrative example I



Illustrative example I

- The sample mean vector for the data is given by:

$$\overline{x} = (4246.42, 4435.80, 1.17, 70.87, 7.37, 53.10, 104.46, 70735.88)'$$

- The sample covariance matrix is given by:

$$S_x = \begin{pmatrix} 19.93 \times 10^6 & 57.12 \times 10^3 & 292.86 & -407.84 & 5663.52 & -3551.50 & -77.08 \times 10^3 & 8.58 \times 10^6 \\ 57.12 \times 10^3 & 37 \times 10^4 & -163.70 & 280.66 & -521.89 & 3076.76 & 7227.60 & 1.90 \times 10^7 \\ 292.86 & -163.70 & 0.37 & -0.48 & 1.58 & -3.23 & -21.29 & 4.01 \times 10^3 \\ -407.84 & 280.66 & -0.48 & 1.80 & -3.86 & 6.31 & 18.28 & -1.22 \times 10^4 \\ 5663.52 & -521.89 & 1.58 & -3.86 & 13.62 & -14.54 & -103.40 & 7.19 \times 10^4 \\ -3551.50 & 3076.76 & -3.23 & 6.31 & -14.54 & 65.23 & 153.99 & 2.29 \times 10^5 \\ -77.08 \times 10^3 & 7227.60 & -21.29 & 1828 & -103.40 & 153.99 & 2702.00 & 2.62 \times 10^5 \\ 85.87 \times 10^5 & 1.90 \times 10^7 & 4.01 \times 10^3 & -1.22 \times 10^4 & 7.19 \times 10^4 & 2.29 \times 10^5 & 2.62 \times 10^5 & 7.28 \times 10^9 \end{pmatrix}$$

Illustrative example I

- The eigenvectors of S_x are the columns of the $V_8^{S_x}$ matrix given by:

$$V_8^{S_x} = \begin{pmatrix} -0.00 & 0.99 & 0.02 & -0.00 & 0.00 & 0.00 & -0.00 & 0.00 \\ -0.00 & 0.02 & -0.99 & 0.02 & -0.00 & -0.00 & 0.00 & 0.00 \\ -0.00 & 0.00 & 0.00 & 0.00 & -0.04 & -0.03 & 0.02 & 0.99 \\ 0.00 & -0.00 & 0.00 & -0.00 & 0.11 & 0.28 & 0.95 & -0.01 \\ -0.00 & 0.00 & 0.00 & 0.02 & -0.23 & -0.92 & 0.30 & -0.04 \\ -0.00 & -0.00 & 0.00 & -0.02 & 0.96 & -0.26 & -0.04 & 0.03 \\ -0.00 & -0.00 & 0.00 & -0.98 & -0.03 & -0.01 & 0.00 & 0.00 \\ -0.99 & -0.00 & 0.00 & -0.00 & -0.00 & 0.00 & -0.00 & -0.00 \end{pmatrix}$$

- Note that the first eigenvector is associated with the last variable which is the one with largest sample variance and so on with the other ones.

Illustrative example I

- The eigenvalues of S_x are $\lambda_1^{S_x} = 7.28 \times 10^9$, $\lambda_2^{S_x} = 1.99 \times 10^7$, $\lambda_3^{S_x} = 3.12 \times 10^5$, $\lambda_4^{S_x} = 2.15 \times 10^3$, $\lambda_5^{S_x} = 36.51$, $\lambda_6^{S_x} = 6.05$, $\lambda_7^{S_x} = 0.43$ and $\lambda_8^{S_x} = 0.08$.

- The proportion of variability explained by the components are 0.997, 2.73×10^{-3}, 4.28×10^{-5}, 2.94×10^{-7}, 5.00×10^{-9}, 8.29×10^{-10}, 5.93×10^{-11} and 1.15×10^{-11}, respectively.

- Consequently, the first proportion of accumulated variability is 0.997, while the others are larger than 0.99999.

Illustrative example I

- We standardize the data matrix.
- Therefore, we obtain eigenvalues and eigenvectors of the sample correlation matrix of the data set which is given by:

$$R_x = \begin{pmatrix} 1 & 0.20 & 0.10 & -0.06 & 0.34 & -0.09 & -0.33 & 0.02 \\ 0.20 & 1 & -0.43 & 0.34 & -0.23 & 0.61 & 0.22 & 0.36 \\ 0.10 & -0.43 & 1 & -0.58 & 0.70 & -0.65 & -0.67 & 0.07 \\ -0.06 & 0.34 & -0.58 & 1 & -0.78 & 0.58 & 0.26 & -0.10 \\ 0.34 & -0.23 & 0.70 & -0.78 & 1 & -0.48 & -0.53 & 0.22 \\ -0.09 & 0.61 & -0.65 & 0.58 & -0.48 & 1 & 0.36 & 0.33 \\ -0.33 & 0.22 & -0.67 & 0.26 & -0.53 & 0.36 & 1 & 0.05 \\ 0.02 & 0.36 & 0.07 & -0.10 & 0.22 & 0.33 & 0.05 & 1 \end{pmatrix}$$

Illustrative example I

- The eigenvectors of R_x are the columns of the $V_8^{R_x}$ matrix given by:

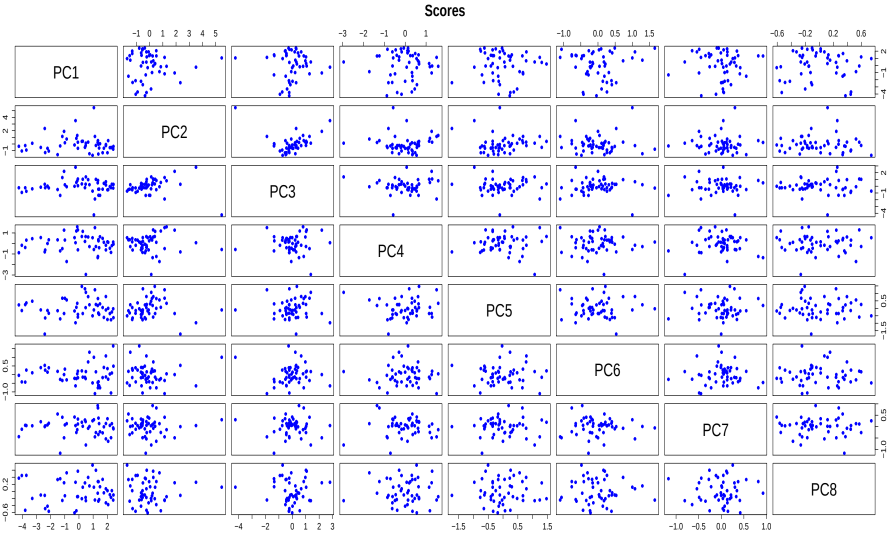
$$V_8^{R_x} = \begin{pmatrix} -0.12 & 0.41 & 0.65 & 0.40 & -0.40 & -0.01 & -0.06 & 0.21 \\ 0.29 & 0.51 & 0.10 & 0.08 & 0.63 & 0.46 & 0.00 & -0.06 \\ -0.46 & 0.05 & -0.07 & -0.35 & -0.00 & 0.38 & -0.61 & 0.33 \\ 0.41 & -0.08 & 0.35 & -0.44 & -0.32 & 0.21 & -0.25 & -0.52 \\ -0.44 & 0.30 & -0.10 & 0.16 & 0.12 & -0.32 & -0.29 & -0.67 \\ 0.42 & 0.29 & -0.04 & -0.23 & 0.09 & -0.64 & -0.39 & 0.30 \\ 0.35 & -0.15 & -0.38 & 0.61 & -0.21 & 0.21 & -0.47 & -0.02 \\ 0.03 & 0.58 & -0.51 & -0.20 & -0.49 & 0.14 & 0.28 & -0.01 \end{pmatrix}$$

- The eigenvalues of R_x are $\lambda_1^{R_x} = 3.59$, $\lambda_2^{R_x} = 1.63$, $\lambda_3^{R_x} = 1.11$, $\lambda_4^{R_x} = 0.70$, $\lambda_5^{R_x} = 0.38$, $\lambda_6^{R_x} = 0.30$, $\lambda_7^{R_x} = 0.14$ and $\lambda_8^{R_x} = 0.11$.

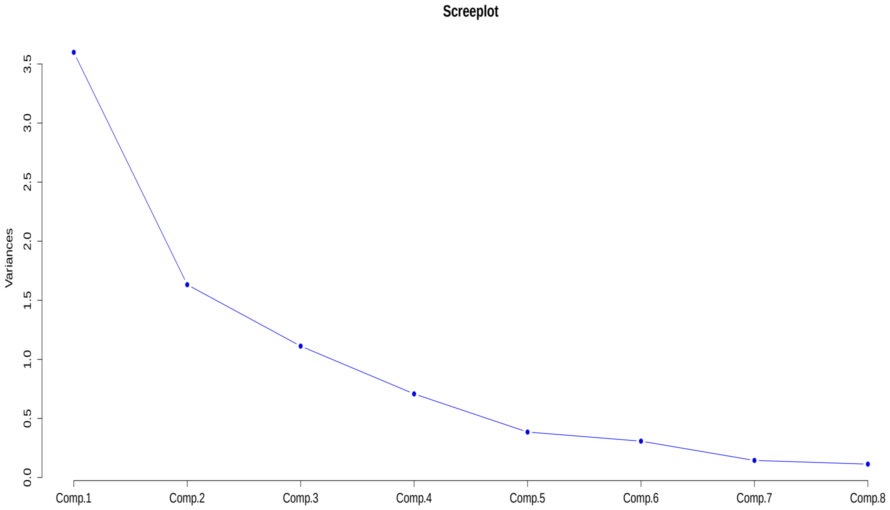
- The proportion of variability explained by the components are 0.449, 0.203, 0.138, 0.088, 0.048, 0.038, 0.018 and 0.014.

- Consequently, the proportion of accumulated variability are 0.449, 0.653, 0.792, 0.881, 0.929, 0.967, 0.985 and 1, respectively.

Illustrative example I



Illustrative example I



Illustrative example I

- In this case, we can select the first three principal components as they explain the 79% of the total variability and the mean of the eigenvalues of the sample covariance or correlation matrix is 1 ($\lambda_3^{R_x} = 1.11$ and $\lambda_4^{R_x} = 0.70$).

- The first component is given by:

$$z_1 = -0.12\widetilde{x}_1 + 0.29\widetilde{x}_2 - 0.46\widetilde{x}_3 + 0.41\widetilde{x}_4 - 0.44\widetilde{x}_5 + 0.42\widetilde{x}_6 + 0.35\widetilde{x}_7 + 0.03\widetilde{x}_8$$

- The first principal component distinguishes between cold states with rich, long-lived, and educated populations, from warm states with poor, short-lived, ill-educated and violent states.

Illustrative example I

- The second component is given by:

$$z_2 = 0.41\widetilde{x}_1 + 0.51\widetilde{x}_2 + 0.05\widetilde{x}_3 - 0.08\widetilde{x}_4 + 0.30\widetilde{x}_5 + 0.29\widetilde{x}_6 - 0.15\widetilde{x}_7 + 0.58\widetilde{x}_8$$

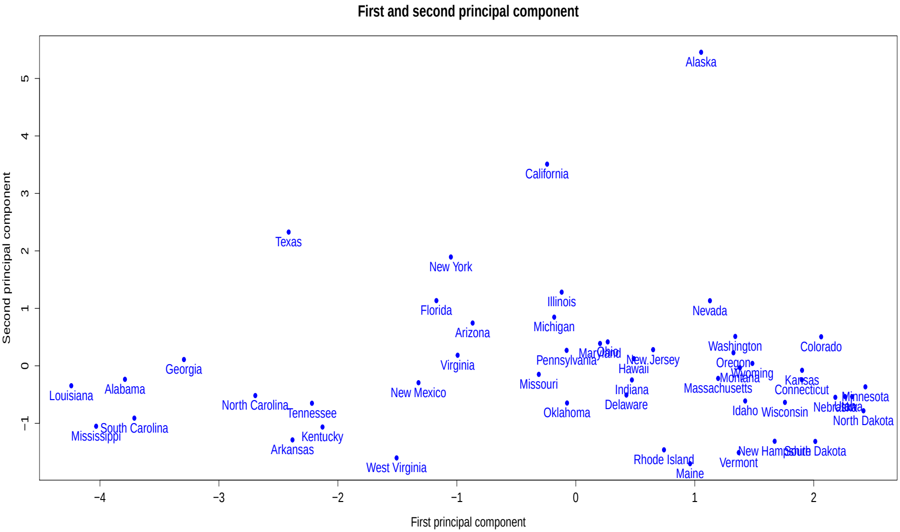
- The second principal component distinguishes big and populated states with rich and educated, although violent states, from small and low populated states with poor and ill-educated people.

- The third component is given by:

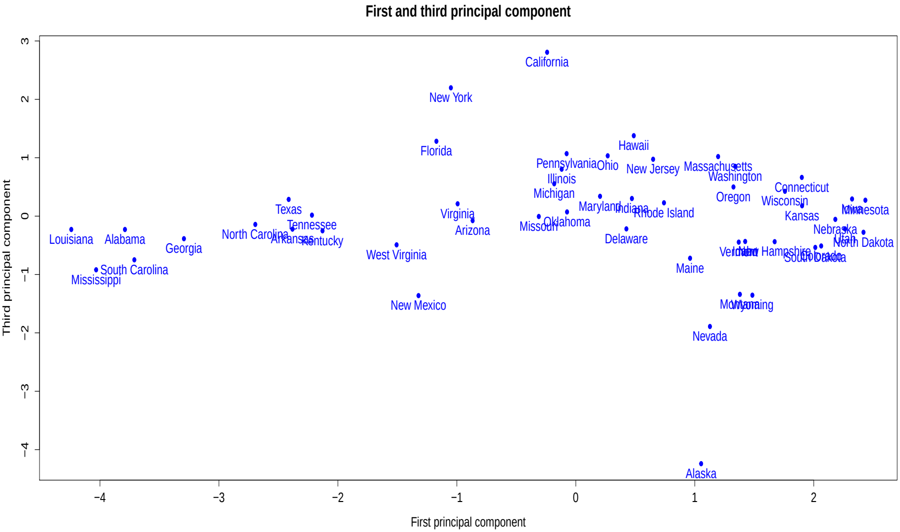
$$z_3 = 0.65\widetilde{x}_1 + 0.10\widetilde{x}_2 - 0.07\widetilde{x}_3 + 0.35\widetilde{x}_4 - 0.10\widetilde{x}_5 - 0.04\widetilde{x}_6 - 0.38\widetilde{x}_7 - 0.51\widetilde{x}_8$$

- The third principal component distinguishes populated states with rich and long-lived populations from warm and big states that tends to be ill-educated and violent.

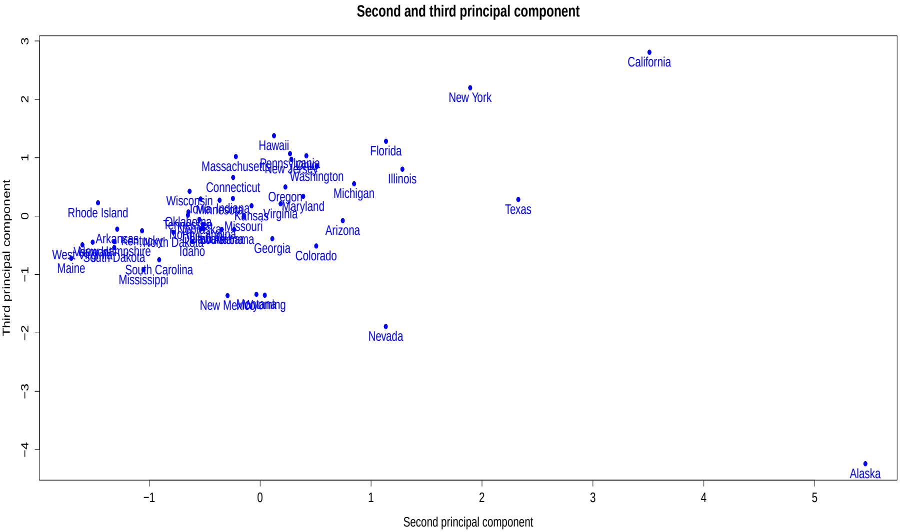
Illustrative example I



Illustrative example I



Illustrative example I



Illustrative example I

- Finally, the correlation between the principal components based on the standardized variables and the original ones are given by:

$$Cor\,(z,x) = \begin{pmatrix} -0.23 & 0.56 & -0.88 & 0.78 & -0.84 & 0.80 & 0.67 & 0.06 \\ 0.52 & 0.66 & 0.06 & -0.10 & 0.39 & 0.38 & -0.19 & 0.75 \\ 0.69 & 0.10 & -0.07 & 0.37 & -0.11 & -0.05 & -0.40 & -0.53 \end{pmatrix}$$

Chapter outline

- We are ready now for:

Chapter 4: Factor Analysis