# Principal Component Analysis

National Chengchi University
Department of Statistics

# Table of Contents


1. What is PCA & Why PCA

2. Theory of PCA

# What is PCA & Why PCA

## Introduction

- A principle component analysis is concerned with explaining the variance-covariance structure of a set of variabe through a few "linear" combinations of these variables
- Objectives of a principle component analysis:
  - data reduction: the total variability of $p$ variables can be accounted for by $k$ principle components, where $p > k$
  - interpretation: can reveal relationship that were not previously suspected

# Theory of PCA

**Population Principal Components**

- Principle components depend solely on the covariance matrix
- Development of principle components does not require a multivariate normal assumption.
- However, a multivariate normal assumption is useful for inference of the principle components.

# Brief Review of Basic Properties

Let $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_p)$ be a random vector and $E(\mathbf{X}) = \boldsymbol{\mu}$, $\text{Cov}(\mathbf{X}) = \boldsymbol{\Sigma}$, then it follows that for any $m \times n$ matrix $A$ and vector $b$ of appropriate size:

- $E(A\mathbf{X} + b) = \boldsymbol{\mu} + b$
- $\text{Cov}(A\mathbf{X} + b) = A\boldsymbol{\Sigma}A' = \text{Cov}(A\mathbf{X})$

Let $\mathbf{a}, \mathbf{b}$ be vectors of appropriate size and $\mathbf{X}, \mathbf{Y}$ be random vectors, then

- $\text{Cov}(\mathbf{a'X}, \mathbf{b'Y}) = \mathbf{a}'\text{Cov}(\mathbf{X,Y})\mathbf{b}$

For any $n \times n$ matrix $A$

- $tr(A) = \sum_{i=1}^{n} \lambda_i$
  where $\lambda_i$'s are the eigenvalues of $A$

- Let $\mathbf{X}^T = [X_1, X_2, \ldots, X_p]$, $\text{Cov}(\mathbf{X}) = \mathbf{\Sigma}$
- $\mathbf{\Sigma}$ has eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$
- Consider

$$
\begin{aligned}
Y_1 &= \mathbf{a}_1^T \mathbf{X} = a_{11} X_1 + a_{12} X_2 + \ldots + a_{1p} X_p \\
Y_2 &= \mathbf{a}_2^T \mathbf{X} = a_{21} X_1 + a_{22} X_2 + \ldots + a_{2p} X_p \\
&\vdots \qquad \vdots \qquad\qquad\qquad \vdots \\
Y_p &= \mathbf{a}_p^T \mathbf{X} = a_{p1} X_1 + a_{p2} X_2 + \ldots + a_{pp} X_p
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow \quad \text{Var}(Y_i) &= \mathbf{a}_i^T \mathbf{\Sigma} \mathbf{a}_i \quad i = 1, \ldots, p \\
\Rightarrow \quad \text{Cov}(Y_i, Y_k) &= \mathbf{a}_i^T \mathbf{\Sigma} \mathbf{a}_k \quad i, k = 1, \ldots, p
\end{aligned}
$$

Define

$$\begin{array}{rcl}
\text{1st principal} & = & \text{linear combination } \mathbf{a}_1^T\mathbf{X} \text{ that maximizes} \\
\text{component} & & \text{Var}(\mathbf{a}_1^T\mathbf{X}) \text{ subject to } \mathbf{a}_1^T\mathbf{a}_1 = 1 \\
\text{2nd principle} & = & \text{linear combination } \mathbf{a}_2^T\mathbf{X} \text{ that maximizes} \\
\text{component} & & \text{Var}(\mathbf{a}_2^T\mathbf{X}) \text{ subject to } \mathbf{a}_2^T\mathbf{a}_2 = 1 \text{ and} \\
& & \text{Cov}(\mathbf{a}_1^T\mathbf{X}, \mathbf{a}_2^T\mathbf{X}) = 0 \\
\vdots & & \vdots \\
i\text{th principle} & = & \text{linear combination } \mathbf{a}_i^T\mathbf{X} \text{ that maximizes} \\
\text{component} & & \text{Var}(\mathbf{a}_i^T\mathbf{X}) \text{ subject to } \mathbf{a}_i^T\mathbf{a}_i = 1 \text{ and} \\
& & \text{Cov}(\mathbf{a}_i^T\mathbf{X}, \mathbf{a}_k^T\mathbf{X}) = 0 \qquad \forall k < i
\end{array}$$

- Results: Let $\boldsymbol{\Sigma}$ be the covariance matrix of $\mathbf{X}^T = [\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_p]$
  Let $\boldsymbol{\Sigma}$ have the eigenvalue-(normalized) vector pairs $(\lambda_1, \mathbf{e}_1), \ldots, (\lambda_p, \mathbf{e}_p)$, where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$ and $\mathbf{e}_i^T \mathbf{e}_k = 1$ if $i = k$; 0 if $i \neq k$.
  Then the $i$th principle component is given by

  $$Y_i = \mathbf{e}_i \mathbf{X} = e_{i1} X_1 + e_{i2} X_2 + \ldots e_{ip} X_p \quad i = 1, \ldots, p$$

  with these choices,

  $$\begin{aligned} \mathrm{Var}(Y_i) &= \mathbf{e}_i^T \boldsymbol{\Sigma} \mathbf{e}_i = \lambda_i \quad & i = 1, \ldots, p \\ \mathrm{Cov}(Y_i, Y_k) &= \mathbf{e}_i^T \boldsymbol{\Sigma} \mathbf{e}_k = 0 \quad & i = 1, \ldots, p \end{aligned}$$

  If some $\lambda_i$ are equal, the choice of $\mathbf{e}_i$, and hence $Y_i$, are not unique.

- Result: Let $\mathbf{X}^T = [X_1, X_2, \ldots, X_p]$ have covariance matrix $\boldsymbol{\Sigma}$ with eigenvalue-eigenvector pairs $(\lambda_1, \mathbf{e}_1), \ldots, (\lambda_p, \mathbf{e}_p)$ where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$
  Let $Y_1 = \mathbf{e}_1^T \mathbf{X}$, $Y_2 = \mathbf{e}_2^T \mathbf{X}, \ldots, Y_p = \mathbf{e}_p^T \mathbf{X}$ be the principle components. Then

$$
\begin{aligned}
\underbrace{\sigma_{11} + \sigma_{22} + \ldots + \sigma_{pp}}_{\text{total population variance}} &= \sum_{i=1}^{p} \mathrm{Var}(X_i) \\
&= \lambda_1 + \lambda_2 + \ldots + \lambda_p \\
&= \sum_{i=1}^{p} \mathrm{Var}(Y_i)
\end{aligned}
$$

- By the above result
  the proportion of total population variance due to the kth principle component equals to

$$\frac{\lambda_k}{\lambda_1 + \lambda_2 + \ldots + \lambda_p} \qquad k = 1, \ldots, p$$

  If most (e.g. 80% to 90%) of the total population variance can be explained by the first $q$ principle components($q < p$), then these $q$ principle components can replace the original $p$ variables without much loss of information (total variance)

- Result: If $Y_1 = \mathbf{e}_1^T\mathbf{X}, \dots, Y_p = \mathbf{e}_p^T\mathbf{X}$ are the principle components obtained from the covariance matrix $\Sigma$, then the correlation coefficient between the component $Y_i$ and the variable $X_k$

$$\rho_{Y_i,X_k} = \frac{e_{ik}\sqrt{\lambda_i}}{\sigma_{kk}} \quad i,k = 1,\dots,p,$$

where $e_{ik}$ is the $k$th component of $\mathbf{e}_i$

Proof.

$\mathrm{Cov}(X_k, Y_i) = \mathrm{Cov}(\mathbf{a}_k\mathbf{X}, \mathbf{e}_i\mathbf{X}) = \mathbf{a}_k'\boldsymbol{\Sigma}\mathbf{e}_i = \mathbf{a}_k'\lambda_i\mathbf{e}_i = \lambda_i e_{ik}$

$$\rho_{Y_i,X_k} = \frac{\mathrm{Cov}(Y_i, X_k)}{\sqrt{\mathrm{Var}(Y_i)}\sqrt{\mathrm{Var}(X_k)}} = \frac{e_{ik}\lambda_i}{\sqrt{\lambda_i}\sigma_k} = \frac{e_{ik}\sqrt{\lambda_i}}{\sigma_{kk}}$$

$\square$

**Principle components obtained from standardized variable**

- Let

$$Z_1 = \frac{X_1 - u_1}{\sqrt{\sigma_{11}}}, \ldots, Z_p = \frac{X_p - u_p}{\sqrt{\sigma_{pp}}}$$

$$\Rightarrow \mathbf{Z} = (\mathbf{V}^{\frac{1}{2}})^{-1}(\mathbf{X} - \boldsymbol{\mu}), \mathrm{Cov}(\mathbf{Z}) = (\mathbf{V}^{\frac{1}{2}})^{-1}\boldsymbol{\Sigma}(\mathbf{V}^{\frac{1}{2}})^{-1} = \boldsymbol{\rho}$$

where

$$\mathbf{V}^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\sigma_{11}} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \sqrt{\sigma_{pp}} \end{bmatrix}$$

- Result: The $i$th principle component of the standardized variable $\mathbf{Z}^T = [\mathbf{Z}_1, \ldots, \mathbf{Z}_p]$ with Cov($\mathbf{Z}$)=$\boldsymbol{\rho}$ is given by

$$Y_i = \mathbf{e}_i^T \mathbf{Z} \qquad i = 1, 2, \ldots, p$$

Moreover,

$$\sum_{i=1}^{p} \text{Var}(Y_i) = \sum_{i=1}^{p} \text{Var}(Z_i) = p$$

and

$$\rho_{Y_i, Z_k} = \mathbf{e}_{ik} \sqrt{\lambda_i} \qquad i, k = 1, 2, \ldots, p$$

In this case, $(\lambda_1, \mathbf{e}_1), \ldots, (\lambda_p, \mathbf{e}_p)$ are eigenvalue-eigenvector pairs for $\rho$ with $\lambda_1 \geq \ldots \geq \lambda_p$