STATS 206
Applied Multivariate Analysis
Lecture 6: Principal Components

Prathapasinghe Dharmawansa

Department of Statistics
Stanford University
Autumn 2013

Agenda

- Population principal components
- Sample principal components

Introduction

- A multivariate response may often have few dominant components accountable for system variability.
- Principal component analysis (PCA):
  - Investigating and identifying the modes of variation (based on the variance-covariance structure)
  - Objectives:
    a. Dimension reduction
    b. Interpretation and insight
  - Facilitating further analysis, such as multiple regression and cluster analysis, etc.

Population Principal Components (i)

Given p random variables $X_1, \ldots, X_p$

- Principal components:
  - Referring to particular linear combinations of $X_1, \ldots, X_p$
  - Representing the selection a new coordinate system obtained by rotating the original system with $X_1, \ldots, X_p$ as the coordinate axes
  - Consequences: the new axes:
    1. Representing directions of maximum variability
    2. Providing a simpler and more parsimonious description of the covariance structure

Population Principal Components (ii)

Given a random vector $\mathbf{X} = [X_1, X_2, \ldots, X_p]'$ with cov. matrix $\mathbf{\Sigma}$;
Let $\lambda_1 \geq \lambda_1 \geq \ldots \geq \lambda_p \geq 0$ be eigenvalues of $\mathbf{\Sigma}$
Consider the linear combinations:

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

From previous analysis:

$$\begin{aligned} \Longrightarrow \qquad \mathrm{Var}(Y_i) \quad & = \mathbf{a}_i'\mathbf{\Sigma}\mathbf{a}_i, \quad i = 1, 2, \ldots, p \\ \mathrm{Cov}(Y_i, Y_k) \quad & = \mathbf{a}_i'\mathbf{\Sigma}\mathbf{a}_k, \quad i, k = 1, 2, \ldots, p \end{aligned}$$

- Principal components refer to: the uncorrelated linear combinations $Y_1, \ldots, Y_p$ with variances as large as possible (see $\mathrm{Var}(Y_i)$)

Population Principal Components (iii)

- Principal components
  - The first principal component $= \mathbf{a}_1'\mathbf{X}$, where $\mathbf{a}_1$ is solution to the optimization problem:

$$\begin{cases} \max & \mathsf{Var}(\mathbf{a}_1'\mathbf{X}) \\ \text{subject to} & \mathbf{a}_1'\mathbf{a}_1 = 1 \end{cases} \quad \text{or} \quad \begin{cases} \max & \mathbf{a}_1'\mathbf{\Sigma}\mathbf{a}_1 \\ \text{subject to} & \mathbf{a}_1'\mathbf{a}_1 = 1 \end{cases}$$

    Note: $\mathbf{a}_1$ restricted to unit length; otherwise the maximum of $\mathsf{Var}(\mathbf{a}_1'\mathbf{X})$ can go unbounded by scaling.

  - The second principal component $= \mathbf{a}_2'\mathbf{X}$, where $\mathbf{a}_2$ is solution to the following problem:

$$\begin{cases} \max & \mathsf{Var}(\mathbf{a}_2'\mathbf{X}) \\ \text{subject to} & \mathbf{a}_2'\mathbf{a}_2 = 1 \\ & \mathsf{Cov}(\mathbf{a}_1'\mathbf{X}, \mathbf{a}_2'\mathbf{X}) = 0 \end{cases} \quad \text{or} \quad \begin{cases} \max & \mathbf{a}_2'\mathbf{\Sigma}\mathbf{a}_2 \\ \text{subject to} & \mathbf{a}_2'\mathbf{a}_2 = 1 \\ & \mathbf{a}_1'\mathbf{\Sigma}\mathbf{a}_2 = 0 \end{cases}$$

Population Principal Components (iv)

Principal components (Cont'd)

– In general, the i-th principal component $= \mathbf{a}_i'\mathbf{X}$, where $\mathbf{a}_i$ comes from solution to the following:

$$\begin{cases} \max & \text{Var}(\mathbf{a}_i'\mathbf{X}) \\ \text{subject to} & \mathbf{a}_i'\mathbf{a}_i = 1 \\ & \text{Cov}(\mathbf{a}_i'\mathbf{X}, \mathbf{a}_k'\mathbf{X}) = 0, \quad \text{for } k < i \end{cases}$$

$$\text{or} \quad \begin{cases} \max & \mathbf{a}_i'\boldsymbol{\Sigma}\mathbf{a}_i \\ \text{subject to} & \mathbf{a}_i'\mathbf{a}_i = 1 \\ & \mathbf{a}_i'\boldsymbol{\Sigma}\mathbf{a}_k = 0, \quad \text{for } k < i \end{cases}$$

Population Principal Components (v)

Principal components (Cont'd)

- Important result:

Let random vector $\mathbf{X} = [X_1, X_2, \ldots, X_p]'$ have cov. matrix $\mathbf{\Sigma}$; $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \ldots, (\lambda_p, \mathbf{e}_p)$: eigenvalue-eigenvector pairs of $\mathbf{\Sigma}$; $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$; then the i-th principal component is

$$Y_i = \mathbf{e}_i'\mathbf{X} = e_{i1}X_1 + e_{i2}X_2 + \ldots + e_{ip}X_p, \quad i = 1, 2, \ldots, p$$

With the above: $\begin{cases} \text{Var}(Y_i) = \mathbf{e}_i'\mathbf{\Sigma}\mathbf{e}_i = \lambda_i, & i = 1, 2, \ldots, p \\ \text{Cov}(Y_i, Y_k) = \mathbf{e}_i'\mathbf{\Sigma}\mathbf{e}_k = 0, & \forall i \neq k \end{cases}$

If some eigenvalues λ_i of $\mathbf{\Sigma}$ are equal, then the choices of the associated coefficient vectors $\mathbf{e}_i$, and hence Y_i, are not unique.

Population Principal Components (vi)

Proof (outline) for the result on previous page based on 1), 2) below

1) $\max\limits_{\mathbf{a}\neq\mathbf{0}} \dfrac{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}}{\mathbf{a}'\mathbf{a}} = \lambda_1$ (Known result; attained when $\mathbf{a} = \mathbf{e}_1$)

$$= \max_{\mathbf{a}\neq\mathbf{0}} \frac{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}}{\|\mathbf{a}\|^2} = \max_{\frac{\mathbf{a}}{\|\mathbf{a}\|}\neq\mathbf{0}} \frac{\mathbf{a}'}{\|\mathbf{a}\|}\boldsymbol{\Sigma}\frac{\mathbf{a}}{\|\mathbf{a}\|} = \max_{\|\mathbf{a}\|=1} \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} = \underbrace{\max_{\mathbf{a}'\mathbf{a}=1} \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}}_{\text{our def. for the 1st PC}}$$

2) $\max\limits_{\substack{\mathbf{a}\neq\mathbf{0}\\ \mathbf{a}\perp\mathbf{e}_k,\forall k<i}} \dfrac{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}}{\mathbf{a}'\mathbf{a}} = \lambda_i, i = 2, \ldots, p$ (Known result; maximizing $\mathbf{a} = \mathbf{e}_i$)

$$= \max_{\substack{\mathbf{a}'\mathbf{a}=1\\ \mathbf{a}\perp\mathbf{e}_k\\ \forall k<i}} \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} \overset{\text{assuming } \lambda_k>0,\forall k<i}{\Longrightarrow} = \max_{\substack{\mathbf{a}'\mathbf{a}=1\\ \mathbf{a}'\lambda_k\mathbf{e}_k=0\\ \forall k<i}} \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} = \underbrace{\max_{\substack{\mathbf{a}'\mathbf{a}=1\\ \mathbf{a}'\boldsymbol{\Sigma}\mathbf{e}_k=0,\forall k<i}} \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}}_{\text{our def. for the } i\text{-th PC}}$$

(details omitted; see the textbook p. 432)

Population Principal Components (vii)

- Clearly, from the previous result, we have the following:

Let random vector $\mathbf{X} = [X_1, X_2, \ldots, X_p]'$ have covariance matrix $\mathbf{\Sigma}$; $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \ldots, (\lambda_p, \mathbf{e}_p)$: the eigenvalue-eigenvector pairs of $\mathbf{\Sigma}$; Here $\lambda_1 \geq \lambda_1 \geq \ldots \geq \lambda_p \geq 0$. Let $Y_i = \mathbf{e}_i'\mathbf{X}, \quad i = 1, 2, \ldots, p$, be the principal components. Then the total population variance is given by:

$$\sum_{i=1}^{p} \text{Var}(X_i) = \sum_{i=1}^{p} \sigma_{ii} = \text{trace}(\mathbf{\Sigma}) = \sum_{i=1}^{p} \lambda_i = \sum_{i=1}^{p} \text{Var}(Y_i)$$

The proportion of total population variance due to the k-th PC is:

$$\frac{\lambda_k}{\lambda_1 + \lambda_2 + \ldots + \lambda_p} = \frac{\lambda_k}{\sum_{i=1}^{p} \lambda_i} \qquad k = 1, 2, \ldots, p$$

Population Principal Components (viii)

ρ_{Y_i,X_k}: correlation coefficient between random variables Y_i, X_k

If $Y_i = \mathbf{e}_i'\mathbf{X}$, $i = 1, 2, \ldots, p$, are the principal components obtained from the covariance matrix $\mathbf{\Sigma}$ of $\mathbf{X} = [X_1, \ldots, X_p]'$, then we have:

$$\rho_{Y_i,X_k} = \frac{e_{ik}\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}} \qquad i, k = 1, 2, \ldots, p$$

$(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \ldots, (\lambda_p, \mathbf{e}_p)$: eigenvalue-eigenvector pairs of $\mathbf{\Sigma}$

- Proof: Let $\mathbf{a}_k = [0, \ldots, 0, 1, 0, \ldots, 0]'$ with 1 at the k-th position. Then $X_k = \mathbf{a}_k'\mathbf{X}$ and $\mathrm{Cov}(X_k, Y_i) = \mathbf{a}_k'\mathbf{\Sigma}\mathbf{e}_i$. Since $\mathbf{\Sigma}\mathbf{e}_i = \lambda_i\mathbf{e}_i$, $\mathrm{Cov}(X_k, Y_i) = \lambda_i\mathbf{a}_k'\mathbf{e}_i = \lambda_i e_{ik}$. Thus,

$$\rho_{Y_i,X_k} \triangleq \frac{\mathrm{Cov}(X_k, Y_i)}{\sqrt{\mathrm{Var}(Y_i)}\sqrt{\mathrm{Var}(X_k)}} = \frac{e_{ik}\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}}; \quad i, k = 1, \ldots, p \quad \blacksquare$$

Population Principal Components (ix)

Principal Components Obtained from Standardized Variables

- Let $\mathbf{X} = [X_1, \ldots, X_p]'$ have mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Consider the standardized variables:

$$Z_i = \frac{(X_i - \mu_i)}{\sqrt{\sigma_{ii}}}, i = 1, \ldots, p$$

In matrix notation:

$$\mathbf{Z} = (\mathbf{V}^{1/2})^{-1}(\mathbf{X} - \boldsymbol{\mu})$$

$$\text{recall : } \mathbf{V}^{1/2} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{\sigma_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\sigma_{pp}} \end{bmatrix}$$

$$\Longrightarrow \quad \text{Cov}(\mathbf{Z}) = (\mathbf{V}^{1/2})^{-1}\boldsymbol{\Sigma}(\mathbf{V}^{1/2})^{-1} = \boldsymbol{\rho} \text{ (correlation matrix)}$$

Population Principal Components (x)

Principal Components Obtained from Standardized Variables

- Based on previous analysis, we have the following result:

> Let standardized variables $\mathbf{Z} = [Z_1, Z_2, \ldots, Z_p]'$ have cov. matrix $\boldsymbol{\rho}$. $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \ldots, (\lambda_p, \mathbf{e}_p)$ are eigenvalue-eigenvector pairs of $\boldsymbol{\rho}$ and $\lambda_1 \geq \lambda_1 \geq \ldots \geq \lambda_p \geq 0$. The i-th principal component of $\mathbf{Z}$ is
>
> $$Y_i = \mathbf{e}_i'\mathbf{Z} = \mathbf{e}_i'(\mathbf{V}^{1/2})^{-1}(\mathbf{X} - \boldsymbol{\mu}) \quad i = 1, 2, \ldots, p$$
>
> $\sum_{i=1}^p \text{Var}(Y_i) = \sum_{i=1}^p \text{Var}(Z_i) = p$; $\rho_{Y_i,Z_k} = e_{ik}\sqrt{\lambda_i}, i, k = 1, \ldots, p$

- Note:
  1. Principal components from $\boldsymbol{\Sigma}$ and from $\boldsymbol{\rho}$ are different (no simple relation)
  2. Principal component analysis depends on scales of $\mathbf{X}$. Standardization is needed when we want to remove the effect of scaling.

Example 1 — Part (1)

Calculating the Population Principal Components

- Let random vector $\mathbf{X} = [X_1, X_2, X_3]'$ have the covariance matrix

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

The eigenvalue-eigenvector pairs can be shown to be:

$$\begin{cases} \lambda_1 = 5.8284 \\ \mathbf{e}_1 = \begin{bmatrix} 0.3827 \\ -0.9239 \\ 0 \end{bmatrix} \end{cases} \quad \begin{cases} \lambda_2 = 2 \\ \mathbf{e}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \end{cases} \quad \begin{cases} \lambda_3 = 0.1716 \\ \mathbf{e}_3 = \begin{bmatrix} 0.9239 \\ 0.3827 \\ 0 \end{bmatrix} \end{cases}$$

Thus, the principal components are given by

$$Y_1 = \mathbf{e}_1'\mathbf{X} = 0.3827X_1 - 0.9239X_2, \quad Y_2 = \mathbf{e}_2'\mathbf{X} = X_3$$

$$Y_3 = \mathbf{e}_3'\mathbf{X} = 0.9239X_1 + 0.3827X_2$$

Example 1 — Part (2)
Calculating the Population Principal Components

(Cont'd)

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

- Results presented earlier can be demonstrated, e.g.,

$$\begin{aligned}
\mathsf{Var}(Y_1) =& \mathsf{Var}(0.3827X_1 - 0.9239X_2) \\
=& (0.3827)^2 \mathsf{Var}(X_1) + (-0.9239)^2 \mathsf{Var}(X_2) \\
& + 2(0.3827)(-0.9239)\mathsf{Cov}(X_1, X_2) \\
=& 5.8284 = \lambda_1 \\
\mathsf{Cov}(Y_1, Y_2) =& \mathsf{Cov}(0.3827X_1 - 0.9239X_2, X_3) \\
=& 0.3827\mathsf{Cov}(X_1, X_3) - 0.9239\mathsf{Cov}(X_2, X_3) = 0
\end{aligned}$$

Example 1 — Part (3)
Calculating the Population Principal Components

(Cont'd)

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \quad \begin{cases} \lambda_1 = 5.8284 \\ \mathbf{e}_1 = \begin{bmatrix} 0.3827 \\ -0.9239 \\ 0 \end{bmatrix} \end{cases} \quad \begin{cases} \lambda_2 = 2 \\ \mathbf{e}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \end{cases} \quad \begin{cases} \lambda_3 = 0.1716 \\ \mathbf{e}_3 = \begin{bmatrix} 0.9239 \\ 0.3827 \\ 0 \end{bmatrix} \end{cases}$$

- Results presented earlier can be demonstrated, e.g.,

$$\sigma_{11} + \sigma_{22} + \sigma_{33} = 1 + 5 + 2 = \lambda_1 + \lambda_2 + \lambda_3 = 5.8284 + 2 + 0.1716$$

Proportion of total variance from the 1st PC $= \frac{\lambda_1}{\sum_{i=1}^{3} \lambda_i} = \frac{5.8284}{8} = 0.7286$

Proportion of total variance from the first two PCs $= \frac{5.8284+2}{8} = 0.9786$

Correlation coefficients, e.g.:

$$\rho_{Y_1,X_2} = \frac{e_{12}\sqrt{\lambda_1}}{\sqrt{\sigma_{22}}} = \frac{(-0.9239)\sqrt{5.8284}}{\sqrt{5}} = -0.9975$$

Example 2 — Part (1)
Principal Components from Σ and ρ are Different

- Given the covariance matrix $\mathbf{\Sigma}$ of $\mathbf{X}$, we derive the correlation matrix $\boldsymbol{\rho}$ (or the covariance matrix of the standardized variables $\mathbf{Z}$)

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & 4 \\ 4 & 100 \end{bmatrix} \Longrightarrow \boldsymbol{\rho} = \begin{bmatrix} 1 & 0.4 \\ 0.4 & 1 \end{bmatrix}$$

- The eigenvalue-eigenvector pairs for $\mathbf{\Sigma}$:

$$\begin{aligned} \lambda_1 &= 100.16, \quad \mathbf{e}_1 = [0.040 \quad 0.999]' \\ \lambda_2 &= 0.84, \quad \mathbf{e}_2 = [0.999 \quad -0.040]' \end{aligned}$$

PCs from $\mathbf{\Sigma}$: $Y_1 = 0.040X_1 + 0.999X_2; \quad Y_2 = 0.999X_1 - 0.040X_2$

The first PC here explains a portion

$$\frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{100.16}{101} = 0.992$$

of the total population variance.

Example 2 — Part (2)

Principal Components from Σ and ρ are Different

- The eigenvalue-eigenvector pairs for Cov$(\mathbf{Z})$ of standardized $\mathbf{Z}$:

$$\text{Cov}(\mathbf{Z}) = \boldsymbol{\rho} = \begin{bmatrix} 1 & 0.4 \\ 0.4 & 1 \end{bmatrix} \Longrightarrow \begin{cases} \lambda_1 = 1.4, & \mathbf{e}_1 = [0.707 \quad 0.707]' \\ \lambda_2 = 0.6, & \mathbf{e}_2 = [0.707 \quad -0.707]' \end{cases}$$

PCs from $\boldsymbol{\rho}$:

$$\begin{aligned} Y_1 = 0.707 Z_1 + 0.707 Z_2 = & \quad 0.707 \left(\frac{X_1 - \mu_1}{\sqrt{1}} \right) + 0.707 \left(\frac{X_2 - \mu_2}{\sqrt{100}} \right) \\ = & \quad 0.707(X_1 - \mu_1) + 0.0707(X_2 - \mu_2) \\ Y_2 = 0.707 Z_1 - 0.707 Z_2 = & \quad 0.707 \left(\frac{X_1 - \mu_1}{\sqrt{1}} \right) - 0.707 \left(\frac{X_2 - \mu_2}{\sqrt{100}} \right) \\ = & \quad 0.707(X_1 - \mu_1) - 0.0707(X_2 - \mu_2) \end{aligned}$$

In contrast, here the first PC explains a portion $\frac{\lambda_1}{p} = \frac{1.4}{2} = 0.7$ of the total (standardized) population variance.

Example 3 — The Multivariate Normal Case

- Recall:
  $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Longrightarrow$ Constant density of $\mathbf{X}$ on the $\boldsymbol{\mu}$-centered ellipsoids

$$(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = c^2$$

  – Axes $\pm\sqrt{\lambda_i}\mathbf{e}_i$; $(\lambda_i, \mathbf{e}_i)$'s: eigenvalue-eigenvector pairs of $\boldsymbol{\Sigma}$
- For convenience, let $\boldsymbol{\mu} = \mathbf{0}$. Then we have

$$c^2 = \mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x} = \frac{1}{\lambda_1}(\underbrace{\mathbf{e}_1'\mathbf{x}}_{y_1})^2 + \frac{1}{\lambda_2}(\underbrace{\mathbf{e}_2'\mathbf{x}}_{y_2})^2 + \ldots + \frac{1}{\lambda_p}(\underbrace{\mathbf{e}_p'\mathbf{x}}_{y_p})^2$$

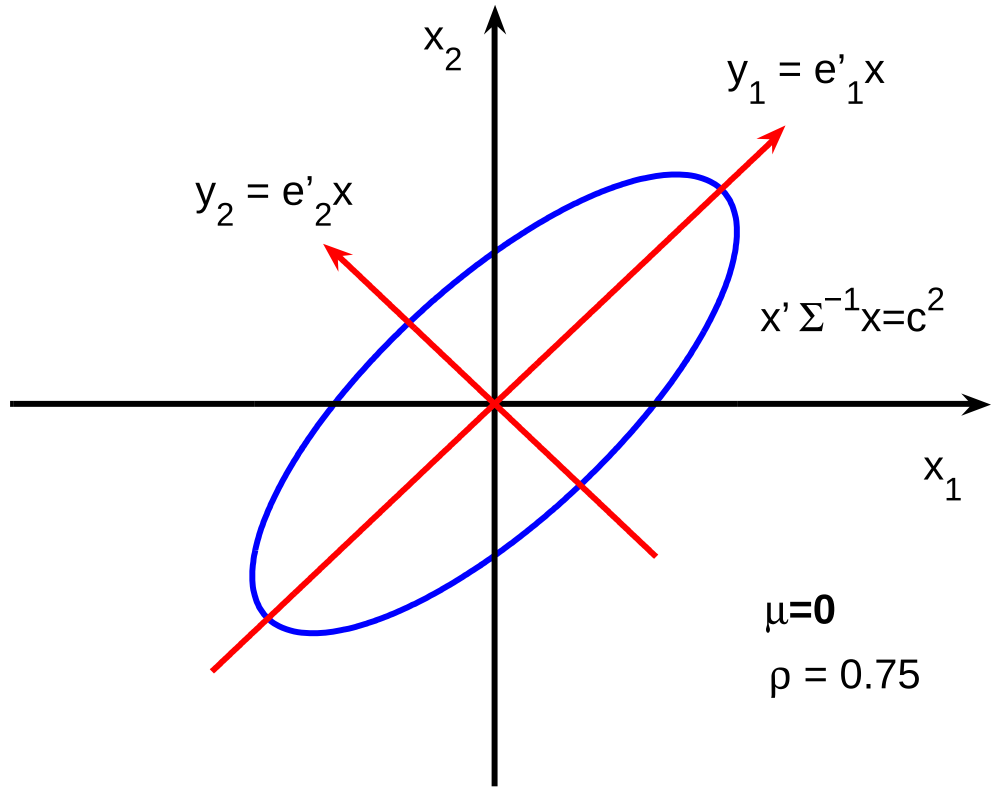
$$= \frac{1}{\lambda_1}y_1^2 + \frac{1}{\lambda_2}y_2^2 + \ldots + \frac{1}{\lambda_p}y_p^2$$

where $y_i = \mathbf{e}_i'\mathbf{x}, i = 1, \ldots, p$, are principal components of $\mathbf{x}$

$\Longrightarrow$Here the PCs lie in the directions of the axes of a constant density ellipsoid!

Example 3 — The Multivariate Normal Case

An Illustration $(p = 2)$



Sample Principal Components (i)

- Let the data $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ be n independent drawings from a p-dimensional population with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

  $\Longrightarrow$ Sample mean $\overline{\mathbf{x}}$, sample cov. matrix $\mathbf{S}$ and sample corr. matrix $\mathbf{R}$

- Objective: to construct uncorrelated linear combinations of the measured characteristics that account for much of the variation in the sample

- Sample principal components: uncorrelated linear combinations with the largest variances

Note: Analyzing the sample PCs is similar to analyzing the population PCs, with $\boldsymbol{\Sigma}$ and $\boldsymbol{\rho}$ replaced by $\mathbf{S}$ and $\mathbf{R}$.

Sample Principal Components (ii)

- Sample principal components:
  - The 1st sample principal component $= \mathbf{a}_1'\mathbf{x}_j$, where $\mathbf{a}_1$ maximizes the sample variance of $\mathbf{a}_1'\mathbf{x}_j$ subject to $\mathbf{a}_1'\mathbf{a}_1 = 1$
  - The 2nd sample principal component $= \mathbf{a}_2'\mathbf{x}_j$, where $\mathbf{a}_2$ maximizes the sample variance of $\mathbf{a}_2'\mathbf{x}_j$ subject to $\mathbf{a}_2'\mathbf{a}_2 = 1$ and zero sample covariance for the pair $(\mathbf{a}_1'\mathbf{x}_j, \mathbf{a}_2'\mathbf{x}_j)$
  - In general, the i-th sample principal component $= \mathbf{a}_i'\mathbf{x}_j$, where $\mathbf{a}_i$ maximizes the sample variance of $\mathbf{a}_i'\mathbf{x}_j$ subject to $\mathbf{a}_i'\mathbf{a}_i = 1$ and zero sample covariances for all pairs $(\mathbf{a}_i'\mathbf{x}_j, \mathbf{a}_k'\mathbf{x}_j), k < i$

Sample Principal Components (iii)

- Important result:

If $\mathbf{S}_{p\times p}$ is the sample cov. matrix with eigenvalue-eigenvector pairs $(\widehat{\lambda}_1, \widehat{\mathbf{e}}_1), (\widehat{\lambda}_2, \widehat{\mathbf{e}}_2), \ldots, (\widehat{\lambda}_p, \widehat{\mathbf{e}}_p)$, then the i-th sample PC is given by:

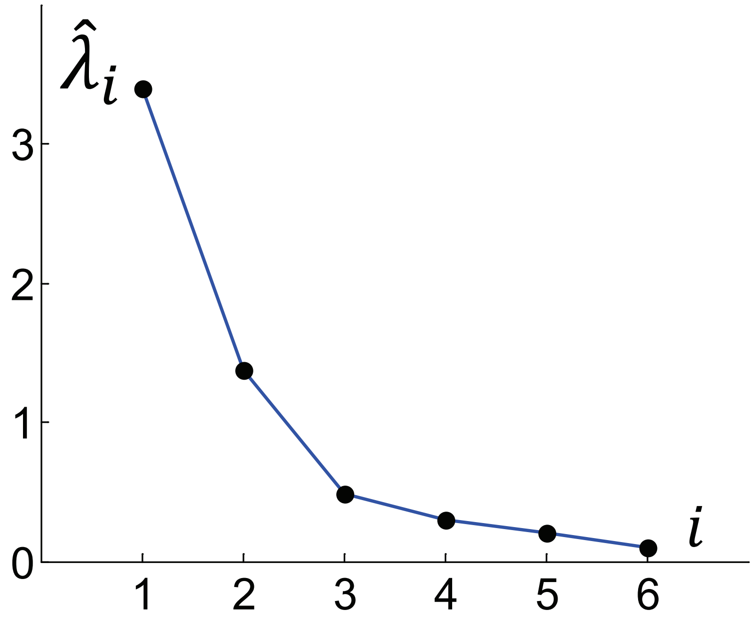
$$\widehat{y}_i = \widehat{\mathbf{e}}_i'\mathbf{x} = \widehat{e}_{i1}x_1 + \widehat{e}_{i2}x_2 + \ldots + \widehat{e}_{ip}x_p, \quad i = 1, 2, \ldots, p$$

where $\widehat{\lambda}_1 \geq \ldots \geq \widehat{\lambda}_p \geq 0$ and $\mathbf{x}$ is any observation on $X_1, \ldots, X_p$.

$$\begin{cases} \text{Sample variance } (\widehat{y}_i) = \widehat{\lambda}_i, & i = 1, \ldots, p \\ \text{Sample covariance } (\widehat{y}_i, \widehat{y}_k) = 0, & \forall i \neq k \\ \text{Total sample variance } = \sum_{i=1}^{p} s_{ii} = \sum_{i=1}^{p} \widehat{\lambda}_i & \\ \text{Correlation coefficient: } r_{\widehat{y}_i, x_k} = \frac{\widehat{e}_{ik}\sqrt{\widehat{\lambda}_i}}{\sqrt{s_{kk}}} & i, k = 1, \ldots, p \end{cases}$$

Sample Principal Components (iv)

How Many Sample Principal Components to Use?



- A visual aid – the scree plot: plotting $\widehat{\lambda}_i$ vs. i, for all i
  Finding the "elbow": here around $i = 3$ (eigenvalues after $\widehat{\lambda}_2$ are small)
- In this case, without any other evidence, 2 (or 3) sample PCs will be effective in summarizing the total sample variance.

Example 4 — Part (1)

Summarizing Sample Variability with One Sample PC

- Data: natural logarithm of $p = 3$ dimensions of $n = 24$ male turtles

Male Turtle Data							
No	Length (ln)	Width (ln)	Height (ln)	**No**	Length (ln)	Width (ln)	Height (ln)
1	4.5326	4.3041	3.6109	**13**	4.7536	4.4998	3.7612
2	4.5433	4.3567	3.5553	**14**	4.7622	4.4998	3.7136
3	4.5643	4.3820	3.5553	**15**	4.7622	4.5109	3.7136
4	4.6151	4.4308	3.6636	**16**	4.7791	4.5326	3.7136
5	4.6250	4.4427	3.6376	**17**	4.7875	4.4886	3.6889
6	4.6347	4.3944	3.6109	**18**	4.7875	4.5326	3.7842
7	4.6444	4.4188	3.6636	**19**	4.7958	4.5539	3.7377
8	4.6634	4.4188	3.6636	**20**	4.8283	4.5326	3.8067
9	4.6728	4.4067	3.6376	**21**	4.8442	4.5643	3.8067
10	4.7185	4.4886	3.6889	**22**	4.8520	4.5539	3.8067
11	4.7274	4.4773	3.6889	**23**	4.8752	4.5539	3.8286
12	4.7362	4.4543	3.6889	**24**	4.9053	4.6634	3.8501

Example 4 — Part (2)

Summarizing Sample Variability with One Sample PC

- Sample mean and cov. matrix from the data:

$$\overline{\mathbf{x}} = \begin{bmatrix} 4.725 \\ 4.778 \\ 3.703 \end{bmatrix}, \qquad \mathbf{S} = 10^{-3} \begin{bmatrix} 11.072 & 8.019 & 8.160 \\ 8.019 & 6.417 & 6.005 \\ 8.160 & 6.005 & 6.773 \end{bmatrix}$$

- Sample principal component analysis: (Based on eigenvalue decomposition of $\mathbf{S}$)

Variable	$\widehat{\mathbf{e}}_1$	$\widehat{\mathbf{e}}_2$	$\widehat{\mathbf{e}}_3$
$\ln$(length)	0.683	-0.159	-0.713
$\ln$(width)	0.510	-0.594	0.622
$\ln$(height)	0.523	0.788	0.324
Variance $\widehat{\lambda}_i$	23.30×10^{-3}	0.60×10^{-3}	0.36×10^{-3}
% of total var.	96.05	2.47	1.48

e.g., from the table, $\widehat{\lambda}_1 = 23.30 \times 10^{-3}$, $\widehat{\mathbf{e}}_1 = [0.683 \;\; 0.510 \;\; 0.523]'$

Example 4 — Part (3)
Summarizing Sample Variability with One Sample PC

(Cont'd)

Variable	$\widehat{\mathbf{e}}_1$	$\widehat{\mathbf{e}}_2$	$\widehat{\mathbf{e}}_3$
$\ln$(length)	0.683	-0.159	-0.713
$\ln$(width)	0.510	-0.594	0.622
$\ln$(height)	0.523	0.788	0.324
Variance $\widehat{\lambda}_i$	23.30×10^{-3}	0.60×10^{-3}	0.36×10^{-3}
% of total var.	96.05	2.47	1.48

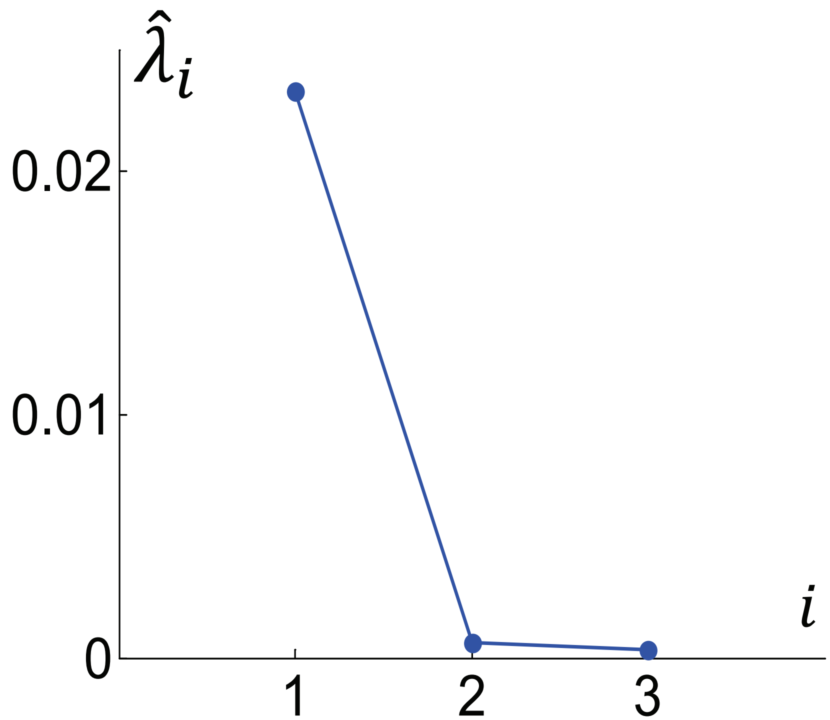
The first principal component is given by:

$$\widehat{y}_1 = 0.683 \ln(\text{length}) + 0.510 \ln(\text{width}) + 0.523 \ln(\text{height})$$

which accounts for 96.05% of the total sample variance.

Example 4 — Part (4)
Summarizing Sample Variability with One Sample PC

- The scree plot for the male turtle data: suggesting one dominant PC is effective in summarizing the total variance



Sample Principal Components (v)

Standardizing the Sample PCs

- Standardization: Construct the following

$$\mathbf{z}_j = (\mathbf{D}^{1/2})^{-1}(\mathbf{x}_j - \overline{\mathbf{x}})$$

$$\mathbf{D}^{1/2} = \begin{bmatrix} \sqrt{s_{11}} & 0 & \dots & 0 \\ 0 & \sqrt{s_{22}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{s_{pp}} \end{bmatrix}$$

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z}_1' \\ \mathbf{z}_2' \\ \vdots \\ \mathbf{z}_n' \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} & \dots & z_{1p} \\ z_{21} & z_{22} & \dots & z_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & \dots & z_{np} \end{bmatrix}$$

Sample Principal Components (vi)
Standardizing the Sample PCs

- It can be shown that the sample mean and covariance matrix are:

$$
\begin{aligned}
\overline{\mathbf{z}} =& \frac{1}{n}\mathbf{Z}'\mathbf{1} = \mathbf{0} \\
\mathbf{S}_z =& \frac{1}{n-1}(\mathbf{Z} - \frac{1}{n}\mathbf{1}\mathbf{1}'\mathbf{Z})'(\mathbf{Z} - \frac{1}{n}\mathbf{1}\mathbf{1}'\mathbf{Z}) \\
=& \frac{1}{n-1}(\mathbf{Z} - \mathbf{1}\overline{\mathbf{z}}')'(\mathbf{Z} - \mathbf{1}\overline{\mathbf{z}}') = \frac{1}{n-1}\mathbf{Z}'\mathbf{Z} \\
=& \begin{bmatrix} 1 & \frac{s_{12}}{\sqrt{s_{11}}\sqrt{s_{22}}} & \cdots & \frac{s_{1p}}{\sqrt{s_{11}}\sqrt{s_{pp}}} \\ \frac{s_{12}}{\sqrt{s_{11}}\sqrt{s_{22}}} & 1 & \cdots & \frac{s_{2p}}{\sqrt{s_{22}}\sqrt{s_{pp}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{s_{1p}}{\sqrt{s_{11}}\sqrt{s_{pp}}} & \frac{s_{2p}}{\sqrt{s_{22}}\sqrt{s_{pp}}} & \cdots & 1 \end{bmatrix} = \mathbf{R}
\end{aligned}
$$

Sample Principal Components (vii)

Standardizing the Sample PCs

- Similarly, we have the following result:

If $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n$ are standardized observations with covariance matrix $\mathbf{R}$, then the i-th sample principal component is given by:

$$\widehat{y}_i = \widehat{\mathbf{e}}_i' \mathbf{z} = \widehat{e}_{i1} z_1 + \widehat{e}_{i2} z_2 + \ldots + \widehat{e}_{ip} z_p, \quad i = 1, 2, \ldots, p$$

$(\widehat{\lambda}_i, \widehat{\mathbf{e}}_i)$: i-th eigenvalue-eigenvector pair of $\mathbf{R}$, $\widehat{\lambda}_1 \geq \ldots \geq \widehat{\lambda}_p \geq 0$

$$\begin{cases} \text{Sample variance } (\widehat{y}_i) = \widehat{\lambda}_i, \quad i = 1, \ldots, p \\ \text{Sample covariance } (\widehat{y}_i, \widehat{y}_k) = 0, \quad \forall i \neq k \\ \text{Total (standardized) sample variance } = \text{trace}(\mathbf{R}) = p = \sum_{i=1}^{p} \widehat{\lambda}_i \\ \text{Correlation coefficient: } r_{\widehat{y}_i, z_k} = \widehat{e}_{ik} \sqrt{\widehat{\lambda}_i}, \quad i, k = 1, \ldots, p \end{cases}$$

Sample Principal Components (viii)

Large Sample Inferences – (1)

Large sample properties of $\widehat{\lambda}_i$ and $\widehat{\mathbf{e}}_i$

- Assumptions here:
  - Observations $\mathbf{X}_1, \ldots, \mathbf{X}_n$: a normal random sample
  - Eigenvalues of the unknown covariance matrix $\mathbf{\Sigma}$ of the normal distribution are positive and distinct: $\lambda_1 > \lambda_2 > \ldots > \lambda_p > 0$
- Results ([Anderson 63][Girshick 39]): Define $\widehat{\boldsymbol{\lambda}} = [\widehat{\lambda}_1, \ldots, \widehat{\lambda}_p]'$

1. Let $\mathbf{\Lambda}$ be the diagonal matrix with entries $\lambda_1, \ldots, \lambda_p$ from $\mathbf{\Sigma}$. Then $\sqrt{n}(\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})$ is approximately $N_p(\mathbf{0}, 2\mathbf{\Lambda})$.
2. Let

$$\mathbf{E}_i \triangleq \lambda_i \sum_{k=1, k\neq i}^{p} \frac{\lambda_k}{(\lambda_k - \lambda_i)^2} \mathbf{e}_k \mathbf{e}_k' \quad \Longrightarrow \quad \sqrt{n}(\widehat{\mathbf{e}}_i - \mathbf{e}_i) \overset{\text{approx.}}{\sim} N_p(\mathbf{0}, \mathbf{E}_i)$$

3. Each $\widehat{\lambda}_i$ is distributed independently of the elements of the associated $\widehat{\mathbf{e}}_i$.

Sample Principal Components (ix)

Large Sample Inferences – (2)

- Result 1 implies that
  - $\widehat{\lambda}_i$'s: independent; each $\widehat{\lambda}_i$: approximately $\sim N(\lambda_i, 2\lambda_i^2/n)$
    $\Longrightarrow P[|\widehat{\lambda}_i - \lambda_i| \leq z(\alpha/2)\lambda_i\sqrt{2/n}] = 1-\alpha$
    A large sample $100(1-\alpha)$% confidence interval for λ_i is:
    $$\frac{\widehat{\lambda}_i}{1+z(\alpha/2)\sqrt{2/n}} \leq \lambda_i \leq \frac{\widehat{\lambda}_i}{1-z(\alpha/2)\sqrt{2/n}}$$
    where $z(\alpha/2)$ is the upper $100(\alpha/2)$th percentile of $N(0,1)$
    Bonferroni-type intervals for m λ_i's: replacing $z(\alpha/2)$ by $z(\alpha/2m)$
- Result 2 $\Longrightarrow \widehat{\mathbf{e}}_i$: normally distributed around $\mathbf{e}_i$ for large samples, $\forall i$
  - Elements of $\widehat{\mathbf{e}}_i$: correlated;
    This correlation depends on n and differences among (unknown) λ_i's
  - Approximate standard errors for $\widehat{e}_{ik}$'s are given by the square roots of the diagonal elements of $\frac{1}{n}\widehat{\mathbf{E}}_i$, where $\widehat{\mathbf{E}}_i$ is derived from $\mathbf{E}_i$ by $\widehat{\lambda}_i \leftarrow \lambda_i$ and $\widehat{\mathbf{e}}_i \leftarrow \mathbf{e}_i$, for all i

Sample Principal Components (x)

Testing for the Equal Correlation Structure

Testing $H_0 : \boldsymbol{\rho} = \underbrace{\boldsymbol{\rho}_0}_{p \times p} = \begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{bmatrix}$ versus $H_1 : \boldsymbol{\rho} \neq \boldsymbol{\rho}_0$

- Motivation of the test: if H_0 holds, then eigenvalues of $\boldsymbol{\Sigma}$ are not distinct, and thus previous results do not apply!
- Lawley's procedure for the test:

  Let

$$\bar{r}_k = \frac{1}{p-1} \sum_{i=1, i \neq k}^{p} r_{ik}, k = 1, 2, \dots, p; \qquad \bar{r} = \frac{2}{p(p-1)} \sum_{k} \sum_{i<k} r_{ik}$$

  $\bar{r}_k$: average off-diagonal elements in the k-th column (or row) of $\mathbf{R}$

  $\bar{r}$: overall average of off-diagonal elements of $\mathbf{R}$

Sample Principal Components (xi)

Testing for the Equal Correlation Structure

Lawley's procedure (Cont'd)

Further let

$$\widehat{\gamma} = \frac{(p-1)^2[1-(1-\overline{r})^2]}{p-(p-2)(1-\overline{r})^2}$$

Then the large sample approximate α-level test is to reject H_0 in favor of H_1 if

$$T = \frac{n-1}{(1-\overline{r})^2}\left[\sum_k \sum_{i<k}(r_{ik}-\overline{r})^2 - \widehat{\gamma}\sum_{k=1}^{p}(\overline{r}_k-\overline{r})^2\right] > \chi^2_{(p+1)(p-2)/2}(\alpha)$$