Chap 4. Principal Components Analysis

History

- First introduced by Karl Pearson (1901) in *Philosophical Magazine* as a procedure for finding lines and planes which best fit a set of points in p-dimensional space. The focus was on geometric optimization.

Basic Idea

- The general objectives are
  - ✓ Dimension reduction
  - ✓ Interpretation of data

Reduce the dimensionality of a data set in which there is a large number of inter-related variables while retaining as much as possible the variation in the original set of variables.

The reduction is achieved by transforming the original variables to a new set of variables, "principal components, that are uncorrelated and ordered such that the first few retains most of the variation present in the data.

Goals & Objectives

- Reduction and summary $\longrightarrow$ data reduction.
- Study the structure of $\mathbf{\Sigma}$ (or $\mathbf{S}$ or $\mathbf{R}$) $\longrightarrow$ Interpretation.

Applications

- Interpretation (study structure)
- Create a new set of variables (a smaller number that are uncorrelated). These can be used in other procedures (e.g., multiple regression).
- Select a sub-set of the original variables to be used in other multivariate procedures.
- Detect outliers or clusters of observations.
- Check multivariate normality assumption (before assuming multivariate normality and analyzing data using procedures that assume multivariate normality.

• Algebraically, principal components are particular uncorrelated linear combinations of the p random variables $X_1, X_2, \cdots, X_p$.

• Geometrically, principal components represent the selection of a new coordinate system obtained by rotating the original system with $X_1, X_2, \cdots, X_p$ as the coordinate axes.

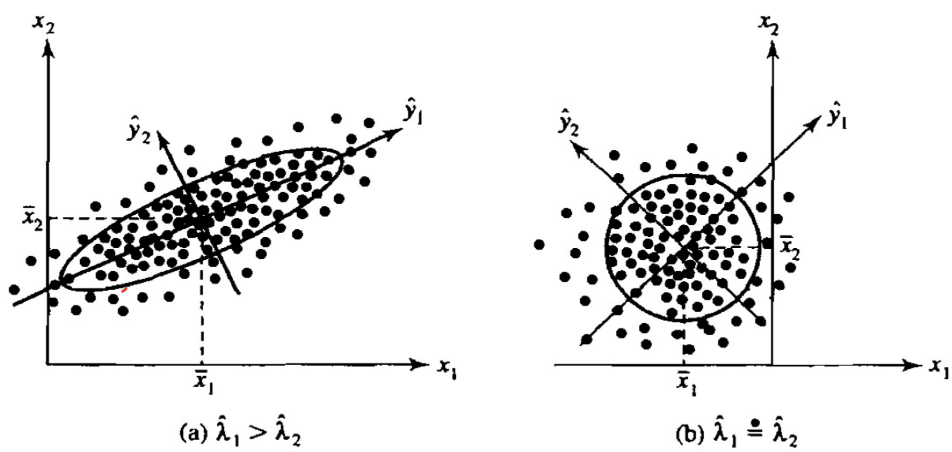


Figure 8.4, Johnson and Wichern (2007)

- PCs represent a selection of a new coordinate system obtained by rotating the original axes to a set of new axes (to provide a simpler structure).
  - The first principal component represents the direction of maximum variability.
  - The second principal component represents the direction of maximum variability that is orthogonal to the first.
  - And so on, until the last PC which represents the direction of minimum variability & orthogonal to all of the others.

4.2 Definition and Derivation of PC's

Matrix Algebra - Revisions

• Principal components represent the directions with maximum variability and provide a simpler and more parsimonious description of the covariance structure.

$\underset{\sim}{X} = \underset{\sim}{Y} - \underset{\sim}{\mu}$

Let the random vector $\mathbf{X} = (X_1, X_2, \cdots, X_p)'$ have the covariance matrix $\mathbf{\Sigma}$ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$, and $E(\underset{\sim}{X}) = \underset{\sim}{0}$

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

$\mathrm{var}(Y_1) =$
$= \mathrm{var}(\underset{\sim}{a}_1^T \underset{\sim}{X}) =$
$= \underset{\sim}{a}_1^T \mathrm{var}(\underset{\sim}{X})\, \underset{\sim}{a}_1 =$
$\underset{\sim}{a}_1^T \underset{\sim}{\Sigma}\, \underset{\sim}{a}_1$

$$\Downarrow$$

$$\mathrm{Var}(Y_i) = \mathbf{a}_i'\mathbf{\Sigma}\mathbf{a}_i \qquad \mathrm{Cov}(Y_i, Y_k) = \mathbf{a}_i'\mathbf{\Sigma}\mathbf{a}_k$$

$\mathrm{var}(\underset{\sim}{a}_i^T \underset{\sim}{X}) = \underset{\sim}{a}_i^T \mathrm{var}(\underset{\sim}{X})\, \underset{\sim}{a}_i$
$= \underset{\sim}{a}_i^T \underset{\sim}{\Sigma}\, a_i$

$\mathrm{cov}(\underset{\sim}{a}_i^T \underset{\sim}{X}, \underset{\sim}{a}_k^T \underset{\sim}{X}) =$
$= \underset{\sim}{a}_i^T \underbrace{\mathrm{cov}(\underset{\sim}{X}, \underset{\sim}{X})}_{\mathrm{var}(\underset{\sim}{X})} \underset{\sim}{a}_k^T = \underset{\sim}{a}_i^T \underset{\sim}{\Sigma}\, \underset{\sim}{a}_k$

Principal components

$$\begin{aligned}
\text{1st PC} &= \text{linear combination } \mathbf{a}_1'\mathbf{X} \text{ that maximizes } \mathrm{Var}(\mathbf{a}_1'\mathbf{X}) \\
&\ \text{s.t.} \quad \mathbf{a}_1'\mathbf{a}_1 = 1 \\
\text{2nd PC} &= \text{linear combination } \mathbf{a}_2'\mathbf{X} \text{ that maximizes } \mathrm{Var}(\mathbf{a}_2'\mathbf{X}) \\
&\ \text{s.t.} \quad \mathbf{a}_2'\mathbf{a}_2 = 1 \text{ and } \mathrm{Cov}(\mathbf{a}_1'\mathbf{X}, \mathbf{a}_2'\mathbf{X}) = 0
\end{aligned}$$

At the ith step,

$$\begin{aligned}
i\text{th PC} &= \text{linear combination } \mathbf{a}_i'\mathbf{X} \text{ that maximizes } \mathrm{Var}(\mathbf{a}_i'\mathbf{X}) \\
&\ \text{s.t.} \quad \mathbf{a}_i'\mathbf{a}_i = 1 \text{ and } \mathrm{Cov}(\mathbf{a}_i'\mathbf{X}, \mathbf{a}_k'\mathbf{X}) = 0, \quad \text{for } k < i.
\end{aligned}$$

$\mathrm{var}(\underset{\sim}{a}\underset{\sim}{X}) = \underset{\sim}{a}^T \underset{\sim}{\Sigma}\, \underset{\sim}{a}$

Principal components - Result 1:

Let the random vector $\mathbf{X} = (X_1, X_2, \cdots, X_p)'$ have the covariance matrix $\mathbf{\Sigma}$ with eigenvalue-eigenvector pairs $(\lambda_1, \underset{\sim}{\gamma}_1)$, $(\lambda_2, \underset{\sim}{\gamma}_2)$, ..., $(\lambda_p, \underset{\sim}{\gamma}_p)$ where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. Then, the *$i$th principal component* is

$$Y_i = \underset{\sim}{\gamma}_i'\mathbf{X} = \gamma_{i1}X_1 + \gamma_{i2}X_2 + \cdots + \gamma_{ip}X_p, \quad i = 1, 2, \ldots, p$$

and

$$\begin{aligned}
&1 \quad \mathrm{Var}(Y_i) = \underset{\sim}{\gamma}_i'\mathbf{\Sigma}\underset{\sim}{\gamma}_i = \lambda_i \text{ for } i = 1, 2, \ldots, p \\
&2 \quad \mathrm{Cov}(Y_i, Y_k) = \underset{\sim}{\gamma}_i'\mathbf{\Sigma}\underset{\sim}{\gamma}_k = 0 \text{ for } i \neq k
\end{aligned}$$

$\mathrm{var}(Y_i) = \mathrm{var}(\underset{\sim}{\gamma}_i^T \underset{\sim}{X}) = \underset{\sim}{\gamma}_i^T \underbrace{\underset{\sim}{\Sigma}\,\underset{\sim}{\gamma}_i}_{\lambda_i \underset{\sim}{\gamma}} = \lambda_i \underbrace{\underset{\sim}{\gamma}_i^T \underset{\sim}{\gamma}_i}_{} = \lambda_i$

Obs: $\underset{\sim}{\gamma}_i^T \underset{\sim}{\gamma}_i = 1$ and $\underset{\sim}{\gamma}_i \underset{\sim}{\gamma}_k = 0$

Norm = 1

$\underset{\sim}{\gamma}_i \perp \underset{\sim}{\gamma}_k$ orthogonal vectors

$\boxed{\underset{\sim}{\Sigma}\,\underset{\sim}{\gamma} = \lambda \underset{\sim}{\gamma}}$

$Y_i = \underset{\sim}{\gamma}_i^T \underset{\sim}{X}$ | $Y_i = \underset{\sim}{\gamma}_i^T (\underset{\sim}{X} - \underset{\sim}{\mu})$

Prove :

$$\begin{aligned}
\frac{\mathbf{a}'\boldsymbol{\Sigma}_X\mathbf{a}}{\mathbf{a}'\mathbf{a}} &= \text{var}(Y_1) \\
\mathbf{a}'\boldsymbol{\Sigma}_X\mathbf{a} &= \text{var}(Y_1)\mathbf{a}'\mathbf{a} \\
\mathbf{a}'\boldsymbol{\Sigma}_X\mathbf{a} - \text{var}(Y_1)\mathbf{a}'\mathbf{a} &= 0 \\
\mathbf{a}'(\boldsymbol{\Sigma}_X\mathbf{a} - \text{var}(Y_1)\mathbf{a}) &= 0 \qquad (\text{since } \mathbf{a} \neq 0) \\
\boldsymbol{\Sigma}_X\mathbf{a} - \text{var}(Y_1)\mathbf{a} &= 0 \\
\underbrace{\boldsymbol{\Sigma}_X}_{p\times p}\underbrace{\mathbf{a}}_{p\times 1} &= \underbrace{\text{var}(Y_1)}_{scalar}\underbrace{\mathbf{a}}_{p\times 1}
\end{aligned}$$

which is just the equation what eigenvalues and eigenvectors solve.

So, $Y_i = \gamma_i^T X$ where γ_i is the i-th eigenvector of Σ.

$cov(Y_i, Y_i) =$

1. $VAR(Y_i) = VAR(\gamma_i^T X) = \gamma_i^T VAR(X)\gamma_i =$

$= \gamma_i^T \underbrace{\Sigma \gamma_i}_{\lambda_i \gamma_i} = \lambda_i \gamma_i^T \gamma_i = \lambda_i$

$$\boxed{\Sigma \gamma = \lambda \gamma}$$

2. $cov(Y_i, Y_k) = cov(\gamma_i^T X, \gamma_k^T X) = \gamma_i' \underbrace{cov(X, X)}\gamma_k =$

$i \neq k$

$\gamma_i^T VAR(X)\gamma_k = \gamma_i^T \underbrace{\Sigma \gamma_k}_{\lambda_k \gamma_k} =$

$= \lambda_k \underbrace{\gamma_i^T \gamma_k} = 0$

Y_i and Y_k are uncorrelated.

$\gamma_i^T \gamma_i = 1$

$\underline{\gamma_i^T \gamma_k = 0}$

Thus, the PC's are uncorrelated and have variance equal to the eigenvalues of the covariance matrix.

Obs:

1. If some eingenvalues are equal the choice of the corresponding eigenvectors are not unique => Pc's is not unique;
2. If $Y_i = \underset{\sim}{\gamma}_i^T \underset{\sim}{X}$ is the i-th P.C. the $Y_i^* = -\underset{\sim}{\gamma}_i^T \underset{\sim}{X}$ is also a P.C. since $(-\underset{\sim}{\gamma}_i^T)(-\underset{\sim}{\gamma}_i) = 1$ and $var(-\underset{\sim}{\gamma}_i^T \underset{\sim}{X}) = var(\underset{\sim}{\gamma}_i^T \underset{\sim}{X}) = \lambda_i$

4.3 Properties of PC+ 4.4 Geometric properties of PC

Principal components - ResulT 2:

Let the random vector $\mathbf{X} = (X_1, X_2, \cdots, X_p)'$ have the covariance matrix $\mathbf{\Sigma}$ with eigenvalue-eigenvector pairs $(\lambda_1, \underset{\sim}{\gamma}_1)$, $(\lambda_2, \underset{\sim}{\gamma}_2)$, ..., $(\lambda_p, \underset{\sim}{\gamma}_p)$ where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. Let $Y_1 = \underset{\sim}{\gamma}_1'\mathbf{X}$, $Y_2 = \underset{\sim}{\gamma}_2'\mathbf{X}$, $\cdots$, $Y_p = \underset{\sim}{\gamma}_p'\mathbf{X}$ be the principal components. Then

$$\sigma_{11} + \cdots + \sigma_{pp} = \sum_{i=1}^{p} \mathrm{Var}(X_i) = \lambda_1 + \cdots + \lambda_p = \sum_{i=1}^{p} \mathrm{Var}(Y_i)$$

Prove:

Matrix Notation:

$$\underset{\sim}{Y} = \underset{\sim}{\Gamma}^T \underset{\sim}{X} \quad \text{where} \quad \underset{\sim}{\Gamma} = \begin{array}{c} \begin{matrix} \underset{\sim}{\gamma}_1 & \underset{\sim}{\gamma}_2 & \cdots & \underset{\sim}{\gamma}_p \end{matrix} \\ \begin{bmatrix} \gamma_{11} & \gamma_{21} & \cdots & \gamma_{p1} \\ \vdots & \vdots & & \vdots \\ \gamma_{1p} & \gamma_{2p} & \cdots & \gamma_{pp} \end{bmatrix} \end{array}$$

$\underset{\sim}{\Gamma}$ matrix: the colums are the eigenvectors of $\underset{\sim}{\Sigma}$ ($\underset{\sim}{\Gamma}$ is an orthogonal matrix)

$Y_1 = \underset{\sim}{\gamma}_1^T \underset{\sim}{X}, \ldots, Y_p = \underset{\sim}{\gamma}_p^T \underset{\sim}{X}$ $\rightarrow \underset{\sim}{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_p \end{bmatrix}$; $\mathrm{var}(Y_i) = \lambda_i$

$\mathrm{cov}(Y_i, Y_j) = 0$

$\underset{\sim}{X} \longrightarrow \mathrm{var}(\underset{\sim}{X}) = \underset{\sim}{\Sigma}$

bigger

$\underset{\sim}{\Lambda} = \mathrm{var}(\underset{\sim}{Y}) = \underbrace{\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & & \\ 0 & \cdots & \cdots & \lambda_p \end{bmatrix}}_{\text{Diagonal}}$

$\underset{\sim}{\Sigma} = \underbrace{\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_2^2 & \cdots & \\ \vdots & & \ddots & \\ & & & \sigma_p^2 \end{bmatrix}}_{\text{not diagonal}}$

$\underbrace{\mathrm{tr}(\underset{\sim}{\Sigma})}_{\text{total variance of } \underset{\sim}{X}} = \sigma_1^2 + \cdots + \sigma_p^2$

$\underbrace{\mathrm{tr}(\underset{\sim}{\Lambda})}_{\text{total variance of } \underset{\sim}{Y}} = \lambda_1 + \cdots + \lambda_p$

total variance of $\underset{\sim}{X} \overset{?}{=}$ total variance of $\underset{\sim}{Y}$

$\sigma_1^2 + \sigma_2^2 + \cdots + \sigma^2 \overset{?}{=} \lambda_1 + \cdots + \lambda_p$

Covariance Matrix:

$\underset{\sim}{\Sigma} = \underbrace{\underset{\sim}{\Gamma}\,\underset{\sim}{\Lambda}\,\underset{\sim}{\Gamma}^T}$

$\downarrow$

Symmetric matrix

$\mathrm{tr}(\underset{\sim}{\Sigma}) = \mathrm{Tr}(\underset{\sim}{\Gamma}\,\underset{\sim}{\Lambda}\,\underset{\sim}{\Gamma}^T) =$

$= \mathrm{Tr}(\underset{\sim}{\Lambda}\,\underbrace{\underset{\sim}{\Gamma}\,\underset{\sim}{\Gamma}^T}_{\underset{\sim}{I}}) = \mathrm{Tr}(\underset{\sim}{\Lambda})$

$\underset{\sim}{\Gamma} = \begin{array}{c} \begin{array}{cccc} \underset{\sim}{\gamma}_1 & \underset{\sim}{\gamma}_2 & \cdots & \underset{\sim}{\gamma}_p \end{array} \\ \begin{bmatrix} \gamma_{11} & & & \gamma_{p1} \\ \gamma_{12} & & & \gamma_{p2} \\ \vdots & & & \vdots \\ \gamma_{1p} & & & \gamma_{pp} \end{bmatrix}_{p \times p} \end{array}$

$\times$ orthogonal matrix : $\underbrace{\underset{\sim}{\Gamma}\,\underset{\sim}{\Gamma}^T = \underset{\sim}{I}}$

$|\underset{\sim}{\Gamma}| = \pm 1$

$$\text{var}(\underset{\sim}{Y}) = \text{var}(\underset{\sim}{T}^T \underset{\sim}{X}) = \underset{\sim}{T}^T \text{var}(\underset{\sim}{X}) \underset{\sim}{T} = \underset{\sim}{T}^T \underset{\sim}{\Sigma} \underset{\sim}{T}$$

using the spectral decomposition of $\underset{\sim}{\Sigma}$

$\underset{\sim}{\Sigma} = \underset{\sim}{T} \underset{\sim}{\Lambda} \underset{\sim}{T}^T$, we have that

$$\text{var}(\underset{\sim}{Y}) = \underset{\sim}{T}^T (\underset{\sim}{T} \underset{\sim}{\Lambda} \underset{\sim}{T}^T) \underset{\sim}{T} = \underbrace{\underset{\sim}{T}^T \underset{\sim}{T}}_{\underset{\sim}{I}} \underset{\sim}{\Lambda} \underbrace{\underset{\sim}{T}^T \underset{\sim}{T}}_{\underset{\sim}{I}} = \underset{\sim}{\Lambda}$$

$$= \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_p \end{bmatrix}$$

$$\text{var}(\underset{\sim}{X}) = \underset{\sim}{\Sigma}$$

total variance of $\underset{\sim}{X} = \text{tr}(\underset{\sim}{\Sigma})$

$$\text{tr}(\underset{\sim}{\Sigma}) = \text{Tr}\begin{bmatrix} \sigma_1^2 & \sigma_{12} \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_2^2 \cdots & \sigma_{2p} \\ \vdots & \vdots & \vdots \\ & & \sigma_p^2 \end{bmatrix} = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_p^2$$

total variance of $\underset{\sim}{Y} = \text{Tr}(\underset{\sim}{\Lambda})$

$$\text{Tr}(\underset{\sim}{\Lambda}) = \lambda_1 + \cdots + \lambda_p$$

But: $\text{Tr}(\underset{\sim}{\Sigma}) = \text{Tr}(\underset{\sim}{T} \underset{\sim}{\Lambda} \underset{\sim}{T}^T) = \text{Tr}(\underbrace{\underset{\sim}{T} \underset{\sim}{T}}_{\underset{\sim}{I}} \underset{\sim}{\Lambda}) =$

$= \text{Tr}(\underset{\sim}{\Lambda}) = \lambda_1 + \cdots + \lambda_p$

Generalized variance of $\underset{\sim}{X} = |\underset{\sim}{\Sigma}|$

$$|\underset{\sim}{\Sigma}| = |\underset{\sim}{\Gamma}\underset{\sim}{\Lambda}\underset{\sim}{\Gamma}^T| \underset{\text{cumutativity}}{=} |\underset{\sim}{\Gamma}||\underset{\sim}{\Lambda}||\underset{\sim}{\Gamma}^T| \underset{|\underset{\sim}{\Gamma}| = \pm 1}{=} |\underset{\sim}{\Lambda}|$$

So, $|\underset{\sim}{\Sigma}| = |\underset{\sim}{\Lambda}| = \lambda_1 \times \lambda_2 \times \cdots \times \lambda_p$

thus, the PC's has the same total and generalized variance as the original variables (X's)

Proportion of total population variance due to kth principal component

$$\frac{\lambda_k}{\underbrace{\lambda_1 + \lambda_2 + \cdots + \lambda_p}_{\text{total variance}}}, \quad k = 1, 2, \ldots, p$$

Remark: If most (for instance, 80 to 90%) of the total population variance, for large p, can be attributed to the first several principal components, then these components can "replace" the original p variables without much loss of information.

• The magnitude of γ_{ik} measures the importance of the kth variable to the ith principal component, irrespective of the other variables.

$$\rho_{Y_i, X_k} = \frac{\gamma_{ik}\sqrt{\lambda_i}}{\sigma_{x_k}}$$

$$\rho_{Y_i, X_k} = \frac{\operatorname{cov}(Y_i, X_k)}{\sigma_{Y_i} \cdot \sigma_{X_k}}$$

$$\underset{\sim}{X} = \begin{bmatrix} X_1 \\ \vdots \\ \boxed{X_k} \\ \vdots \\ X_p \end{bmatrix}$$

$$\operatorname{cov}(Y_i, X_k)$$

$$\underset{\sim}{\gamma}_i^T \underset{\sim}{X}$$

$$\underset{\sim}{e}_k = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leftarrow \text{K-position}$$

$$\underset{\sim}{e}_k^T \underset{\sim}{X} = X_k$$

$$\operatorname{cov}(Y_i, X_k) = \operatorname{cov}(\underset{\sim}{\gamma}_i^T \underset{\sim}{X}, \underset{\sim}{e}_k^T \underset{\sim}{X}) = \underbrace{\underset{\sim}{\gamma}_i^T}_{1\times p} \underbrace{\operatorname{cov}(\underset{\sim}{X}, \underset{\sim}{X})}_{p\times p} \underbrace{\underset{\sim}{e}_k}_{p\times 1}$$

$$\underset{\sim}{\gamma}_{p\times 1}$$

$$\underset{\sim}{\gamma}^T_{1\times p}$$

$$= \underset{\sim}{\gamma}_i^T \underset{\sim}{\Sigma}\, \underset{\sim}{e}_k = \underset{\sim}{e}_k^T \underbrace{\underset{\sim}{\Sigma}\, \underset{\sim}{\gamma}_i}_{\lambda_i \gamma_i} =$$

$$= \lambda_i \underbrace{\underset{\sim}{e}_k^T \underset{\sim}{\gamma}_i}_{\text{K-component of vector } \gamma_i} = \lambda_i \gamma_{ik}$$

$$\frac{\mathrm{cov}(Y_i, X_k)}{\sigma_{X_k}\sigma_{Y_i}} = \frac{\lambda_i \gamma_{ik}}{\sigma_{X_k}\sqrt{\lambda_i}} = \frac{\gamma_{ik}\sqrt{\lambda_i}}{\sigma_k}$$

Principal components – ResulT 3:

Let $Y_1 = \underset{\sim}{\gamma}_1' \mathbf{X}$, $Y_2 = \underset{\sim}{\gamma}_2' \mathbf{X}$, $\cdots$, $Y_p = \underset{\sim}{\gamma}_p' \mathbf{X}$ be the principal components obtained from the covariance matrix $\mathbf{\Sigma}$. Then

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

Prove:

LeT $\underset{\sim}{e}_k^T = (0, 0, \ldots, 1, 0, \ldots, 0)$

k-position

$X_k = \underset{\sim}{e}_k^T \underset{\sim}{X}$; $Y_i = \underset{\sim}{\gamma}_i^T \underset{\sim}{X}$

$$\text{cov}(\underset{\sim}{X}_k, Y_i) = \text{cov}(\underset{\sim}{e}_k^T \underset{\sim}{X}, \underset{\sim}{\gamma}_i^T \underset{\sim}{X}) =$$

$$= \underset{\sim}{e}_k^T \text{cov}(\underset{\sim}{X}, \underset{\sim}{X}) \underset{\sim}{\gamma}_i = \underset{\sim}{e}_k^T \underbrace{\underset{\sim}{\Sigma} \underset{\sim}{\gamma}_i}_{\lambda_i \underset{\sim}{\gamma}_i} =$$

$$= \lambda_i \underset{\sim}{e}_k^T \underset{\sim}{\gamma}_i = \lambda_i \gamma_{ik}$$

$$\rho_{Y_i, X_k} = \frac{\text{cov}(X_k, Y_i)}{\sigma_{X_k} \sigma_{Y_i}} = \frac{\lambda_i \gamma_{ik}}{\sigma_k \sqrt{\lambda_i}} = \frac{\gamma_{ik}\sqrt{\lambda_i}}{\sigma_k}$$

$i, k = 1, \ldots, p$

• The coefficients γ_{ik} and the correlations ρ_{Y_i,X_k} can lead to different rankings as the measures of the importance of the variables to a given component. However, these rankings are often not appreciably different.

• In practice, variables with relatively large coefficients (in absolute value) tend to have relatively large correlations.

• It is suggested that both the coefficients and the correlations be examined to help interpret the principal components.

Example 1: Let $\underset{\sim}{X} \in \mathbb{R}^2$ with $Var(\underset{\sim}{X}) = \underset{\sim}{\Sigma} = \begin{bmatrix} 6 & 2 \\ 2 & 3 \end{bmatrix}$

obtain the Principal components. $E(\underset{\sim}{X}) = \underset{\sim}{\mu}$

$\underset{\sim}{Y} = \underset{\sim}{\Gamma}^T \underset{\sim}{X} \Rightarrow$ obtain the eigenvalues and eigenvectors of $\underset{\sim}{\Sigma}$

1. Eigenvalues:

solution of $|\underset{\sim}{\Sigma} - \lambda \underset{\sim}{I}| = 0 \Leftrightarrow \begin{vmatrix} 6-\lambda & 2 \\ 2 & 3-\lambda \end{vmatrix} = 0 \Leftrightarrow$

$(6-\lambda)(3-\lambda) - 4 = 0 \Leftrightarrow 18 - 6\lambda - 3\lambda + \lambda^2 - 4 = 0 \Leftrightarrow$

$\lambda^2 - 9\lambda + 14 = 0 \Leftrightarrow \lambda = \frac{9 \pm \sqrt{81-56}}{2} \Leftrightarrow \lambda_1 = \frac{9+5}{2} \vee \lambda_2 = \frac{9-5}{2}$

$\Leftrightarrow \lambda_1 = 7$ and $\lambda_2 = 2$

2. Eigenvectors

1st: $\underset{\sim}{\Sigma} \underset{\sim}{x}_1 = \lambda_1 \underset{\sim}{x} \Leftrightarrow \begin{bmatrix} 6 & 2 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} x_{11} \\ x_{12} \end{bmatrix} = 7 \begin{bmatrix} x_{11} \\ x_{12} \end{bmatrix} \Leftrightarrow$

$\begin{cases} 6x_{11} + 2x_{12} = 7x_{11} \\ 2x_{11} + 3x_{12} = 7x_{12} \end{cases} \Leftrightarrow \begin{cases} x_{11} = 2x_{12} \end{cases}$

Select:
$x_{12} = 1 \wedge x_{11} = 2$

$\underset{\sim}{x}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ $\quad \|\underset{\sim}{x}_1\| = \sqrt{2^2+1^2} = \sqrt{5}$ $\therefore \underset{\sim}{\gamma}_1 = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}$ $\quad \underset{\sim}{\gamma}_1 = \frac{\underset{\sim}{x}_1}{\|\underset{\sim}{x}_1\|}$

2nd : $\underset{\sim}{\Sigma}\, \underset{\sim}{x}_2 = \lambda_2 \underset{\sim}{x}_2 \Leftrightarrow$

$$\begin{bmatrix} 6 & 2 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} x_{21} \\ x_{22} \end{bmatrix} = 2 \begin{bmatrix} x_{21} \\ x_{22} \end{bmatrix} \Leftrightarrow \begin{cases} 6x_{21} + 2x_{22} = 2x_{21} \\ 2x_{21} + 3x_{22} = 2x_{22} \end{cases} \Leftrightarrow$$

$$\begin{cases} 4x_{21} = -2x_{22} \end{cases} \Leftrightarrow x_{21} = \tfrac{1}{2} x_{22} \qquad \text{select} : x_{22} = 1 \wedge x_{21} = -\tfrac{1}{2}$$

$$\begin{bmatrix} x_{21} \\ x_{22} \end{bmatrix} = \begin{bmatrix} -1/2 \\ 1 \end{bmatrix} \qquad \|\underset{\sim}{x}_2\| = \sqrt{(-1/2)^2 + 1^2} = \sqrt{\tfrac{5}{4}} = \frac{\sqrt{5}}{2}$$

$$\underset{\sim}{\gamma}_2 = \begin{bmatrix} \dfrac{-1/2}{\sqrt{5}/2} \\ \dfrac{1}{\frac{\sqrt{5}}{2}} \end{bmatrix} = \begin{bmatrix} -\dfrac{1}{\sqrt{5}} \\ \dfrac{2}{\sqrt{5}} \end{bmatrix} \gamma_{22} \qquad \underset{\sim}{\gamma}_2^T \underset{\sim}{\gamma}_1 = 0$$

$$Y_1 = \frac{2}{\sqrt{5}}(X_1 - \mu_1) + \frac{1}{\sqrt{5}}(X_2 - \mu_2) \quad , \operatorname{var}(Y_1) = \lambda_1 = 7$$

$$Y_2 = -\frac{1}{\sqrt{5}}(X_1 - \mu_1) + \frac{2}{\sqrt{5}}(X_2 - \mu_2) \quad , \operatorname{var}(Y_2) = \lambda_2 = 2$$

sum = 9

Obs: $E(Y_1) = E\left[\left(\frac{2}{\sqrt{5}} X_1 - \frac{2}{\sqrt{5}}\mu_1\right) + \left(\frac{1}{\sqrt{5}} X_2 - \frac{1}{\sqrt{5}}\mu_2\right)\right]$

$= 0$ and $E(Y_2) = 0$

IF $E(\underset{\sim}{X}) = \underset{\sim}{0}$ then $\begin{cases} Y_1 = \frac{2}{\sqrt{5}} X_1 + \frac{1}{\sqrt{5}} X_2 & \text{with } E(Y_1) = 0 \\ Y_2 = -\frac{1}{\sqrt{5}} X_1 + \frac{2}{\sqrt{5}} X_2 & \text{with } E(Y_2) = 0 \end{cases}$

Example 2: $\underset{\sim}{X} \in \mathbb{R}^3$ with $E(\underset{\sim}{X}) = \underset{\sim}{0}$; $\underset{\sim}{\Sigma} = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

with eigen values-vectors :

$\lambda_1 = 5.83$; $\underset{\sim 1}{\gamma}^T = [0.383, -0.924, 0]$

$\lambda_2 = 2.0$; $\underset{\sim 2}{\gamma}^T = [0, 0, 1]$

$\lambda_3 = 0.17$; $\underset{\sim 3}{\gamma}^T = [0.924, 0.383, 0]$

$\underset{\sim}{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$

the P.C's are :

$Y_1 = \underset{\sim 1}{\gamma}^T \underset{\sim}{X} = 0.383 X_1 - 0.924 X_2$

$Y_2 = \underset{\sim 2}{\gamma}^T \underset{\sim}{X} = X_3$

$Y_3 = \underset{\sim 3}{\gamma}^T \underset{\sim}{X} = 0.924 X_1 + 0.383 X_2$

total variance of $\underset{\sim}{X}$

$= \sigma_1^2 + \sigma_2^2 + \sigma_3^3 =$

$tr(\underset{\sim}{\Sigma}) = 8$

the Random variable X_3 is one of the P.C's because this variable is not correlated with X_1 and with X_2

$var(aX_1 + bX_2) =$
$a^2 var(X_1) + b^2 var(X_2) + 2ab\, cov(X_1, X_2)$

verify that $var(\underset{\sim}{Y}) = \underset{\sim}{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \\ & \lambda_2 & \\ & & \lambda_3 \end{bmatrix}$

$var(Y_1) = var(0.383 X_1 - 0.924 X_2) = 0.383^2 var(X_1) + 0.924^2 var(X_2) -$

$- 2 \times 0.383 \times 0.924\, cov(X_1, X_2) =$

$= 0.147 \times 1 + 0.854 \times 5 - 0.708(-2) = 5.83 = \lambda_1$

$cov(Y_1, Y_2) = cov(0.383 X_1 - 0.924 X_2, X_3) = 0.383\, cov(X_1, X_3) -$

$- 0.924\, cov(X_2, X_3) = 0.383 \times 0 - 0.924 \times 0 = 0$

$\vdots$

$cov(aX_1 + bX_2, X_3) = cov(aX_1, X_3) + cov(bX_2, X_3)$

$= a\, cov(X_1, X_3) + b\, cov(X_2, X_3)$

variance of $\underset{\sim}{Y}$:

$var(\underset{\sim}{Y}) = \underset{\sim}{\Lambda} = \begin{bmatrix} 5{,}83 & 0 & 0 \\ 0 & 2.0 & 0 \\ 0 & 0 & 0.17 \end{bmatrix}$ $\quad tr(\underset{\sim}{\Lambda}) =$

total variance of $\underset{\sim}{X}$ is the same of the total variance of $\underset{\sim}{Y}$

$$tr(\underset{\sim}{\Sigma}) = \sigma_1^2 + \sigma_2^2 + \sigma_3^2 = 1+5+2 = \lambda_1+\lambda_2+\lambda_3 = 5.83+2+0.17 = 8$$

i	% of total variance explained by the i-th P.C.
1	$\frac{5.83}{8} \times 100\% \simeq 72.9\%$
2	$\frac{2.0}{8} \times 100\% \simeq 25\%$
3	$\frac{0.17}{8} \times 100\% \simeq 2.12\%$

both explained 97.9% of the total variance of $\underset{\sim}{X}$

Correlation between Y_1 and Y_2 with the original variables (X's)

$$\rho_{Y_1,X_1} = \frac{\gamma_{11}\sqrt{\lambda_1}}{\sigma_1} = \frac{0.383\sqrt{5.83}}{1} = 0.925$$

$$\rho_{Y_1,X_2} = \frac{\gamma_{12}\sqrt{\lambda_1}}{\sigma_2} = \frac{-0.924\sqrt{5.83}}{\sqrt{5}} = -0.998$$

$$\rho_{Y_2,X_1} = \rho_{Y_2,X_2} = 0 \quad \text{and} \quad \rho_{Y_2,X_3} = \frac{\sqrt{2}}{\sqrt{2}} = 1$$

obs: the variable X_2 is the one with greater contribution to the 1st PC and also is the one more correlated with Y_1.

Principal Component Analysis

Principal components under normality - Geometric Interpretation

Suppose that $\mathbf{X} \sim N_p(\mathbf{0}, \mathbf{\Sigma})$. Then, $\mathbf{x}'\mathbf{\Sigma}^{-1}\mathbf{x} = c^2$ is an origin-centered ellipsoid which has axes $\pm c\sqrt{\lambda_i}\underset{\sim}{\gamma}_i$, $i = 1, 2, \ldots, p$, where the $(\lambda_i, \underset{\sim}{\gamma}_i)$ are the eigenvalue-eigenvector pairs of $\mathbf{\Sigma}$. Moreover,

$$\begin{aligned} c^2 = \mathbf{x}'\mathbf{\Sigma}^{-1}\mathbf{x} &= \frac{1}{\lambda_1}(\underset{\sim}{\gamma}_1'\mathbf{x})^2 + \frac{1}{\lambda_2}(\underset{\sim}{\gamma}_2'\mathbf{x})^2 + \cdots + \frac{1}{\lambda_p}(\underset{\sim}{\gamma}_p'\mathbf{x})^2 \\ &= \frac{1}{\lambda_1}y_1^2 + \frac{1}{\lambda_2}y_2^2 + \cdots + \frac{1}{\lambda_p}y_p^2 \end{aligned}$$

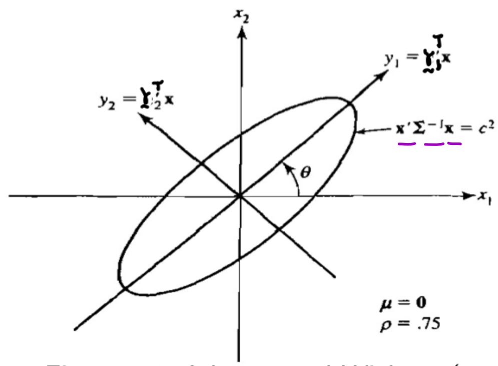


Figure 8.1, Johnson and Wichern (2007)

the PC's Lie in the directions of the axes of the constant density ellipsoid.

$\underset{\sim}{X}$: $E(\underset{\sim}{X}) = \underset{\sim}{0}$

$var(\underset{\sim}{X}) = \underset{\sim}{\Sigma}$

Principal components Scores :

objects marks in the P.C. axes

i-th object observations: $(x_{i1}, x_{i2}, \ldots, x_{ip})$

i-th object on the j-th P.C.

$y_{ij} = \gamma_{j1}x_{i1} + \gamma_{j2}x_{i2} + \cdots + \gamma_{jp}x_{ip}$

score of the i-th object in the j-th P.C

Principal Component Analysis

Standardized principal components

Intead of using $\underset{\sim}{X}^T = (x_1, \dots, x_p)$ and $\operatorname{var}(\underset{\sim}{X}) = \underset{\sim}{\Sigma}$

we can calculate P.C. from $\underset{\sim}{z} = (z_1, \dots, z_p)$, where

$$z_i = \frac{x_i - \mu_i}{\sigma_i}, \; i = 1, \dots, p$$

$\operatorname{var}(\underset{\sim}{z}) = \underset{\sim}{\rho}$

$\underset{\sim}{\rho} = \begin{bmatrix} 1 & - & - & - \\ = & 1 & & \\ = & & \ddots & \\ & & & 1 \end{bmatrix}$ $\operatorname{tr}(\underset{\sim}{\rho}) = p$

Principal Component Analysis

Standardized principal components

$\underset{\sim}{V}^{1/2} = \operatorname{diag}(\sigma_1, \dots, \sigma_p)$

Let $\mathbf{Z} = (V^{1/2})^{-1}(\mathbf{X} - \boldsymbol{\mu})$. Then $E(\mathbf{Z}) = \mathbf{0}$ and

$$\operatorname{Cov}(\mathbf{Z}) = (V^{1/2})^{-1}\boldsymbol{\Sigma}(V^{1/2})^{-1} = \boldsymbol{\rho}$$

The ith principal component of $\mathbf{Z} = (Z_1, Z_2, \cdots, Z_p)'$ is given by

$$Y_i = \underset{\sim}{\gamma}_i'\mathbf{Z} = \underset{\sim}{\gamma}_i'(V^{1/2})^{-1}(\mathbf{X} - \boldsymbol{\mu})$$

and

$$\sum_{i=1}^p \operatorname{Var}(Y_i) = \sum_{i=1}^p \operatorname{Var}(Z_i) = p, \qquad \rho_{Y_i, Z_k} = \gamma_{ik}\sqrt{\lambda_i}$$

where the $(\lambda_i, \underset{\sim}{\gamma}_i)$ are the eigenvalue-eigenvector pairs of $\boldsymbol{\rho}$

Remark: The eigenvalue-eigenvector pairs derived from $\boldsymbol{\Sigma}$ are, in general, not the same as the ones derived from $\boldsymbol{\rho}$.

$\boxed{\text{The PCs from } \boldsymbol{\Sigma}_X \text{ are \underline{not} the same as PCs from } \underset{\sim}{\rho}}$

We'll look at a situation where standardization makes a difference

This will be the case when the scales of the X variables are (substantially or vastly) different and they are ont comparable.

Example 3: Consider the $\underset{\sim}{X} \in \mathbb{R}^2$; $E(\underset{\sim}{X}) = \underset{\sim}{0}$

covariance Matrix and the correlation Matrix

$$\underset{\sim}{\Sigma} = \begin{bmatrix} 1 & 4 \\ 4 & 100 \end{bmatrix} \qquad \underset{\sim}{\rho} = \begin{bmatrix} 1 & 0.4 \\ 0.4 & 1 \end{bmatrix}$$

the eigenvalue - eigenvector pairs from

$$\underset{\sim}{\Sigma} : \begin{cases} \lambda_1 = 100.16 ; & \underset{\sim}{\gamma}_1^T = [0.040; \ 0.999] \\ \lambda_2 = 0.84 ; & \underset{\sim}{\gamma}_2 = [0.999; -0.040] \end{cases}$$

$$\underset{\sim}{\rho} : \begin{cases} \lambda_1^* = 1.4 ; & \underset{\sim}{\gamma}_1^{*T} = [0.707 ; 0.707] \\ \lambda_2^* = 0.6 ; & \underset{\sim}{\gamma}_2^{*T} = [0.707 ; -0.707] \end{cases}$$

$$Z_1 = \frac{X_1 - 0}{\sigma_1}$$

the P.C's are using

$$\underset{\sim}{\Sigma} : \begin{cases} Y_1 = 0.040 \ X_1 + 0.999 \ X_2 \\ Y_2 = 0.999 \ X_1 - 0.040 X_2 \end{cases} \qquad \rho : \begin{cases} Y_1^* = 0.707 \ Z_1 + 0.707 \ Z_2 \\ Y_2^* = 0.707 Z_1 - 0.707 \ Z_2 \end{cases}$$

- we see that X_2 completely dominates the 1st P.C. of $\underset{\sim}{\Sigma}$ this component Y_1 explains $\frac{\lambda_1}{\lambda_1+\lambda_2} = \frac{100.16}{101} = 0.992$ of the total variance of $\underset{\sim}{X}$
- In contrast, the variables Z_1 and Z_2 contribute equally to the P.C's of $\underset{\sim}{\rho}$

In this case the 1st P.C. explains $\frac{\lambda_1}{p} = \frac{1.4}{2} = 0.7$ of the total variance of the standardized variables $(\underset{\sim}{Z})$

$$\rho_{Y_1, X_1} = \frac{\gamma_{11}\sqrt{\lambda_1}}{\sigma_1} = 0.04\sqrt{100.16} \simeq 0.4$$

$$\rho_{Y_1, X_2} = \frac{\gamma_{12}\sqrt{\lambda_1}}{\sigma_2} = \frac{0.999\sqrt{100.16}}{10} \simeq 0.999$$

$$\rho_{Y_1^*, Z_1} = \gamma_{11}^*\sqrt{\lambda_1^*} = 0.707\sqrt{1.4} = 0.837$$

$$\rho_{Y_1^*, Z_2} = \gamma_{12}^*\sqrt{\lambda_1^*} = 0.837$$

we conclude that the relative importance of the variables is affected by standardization.

conclusion: the P.C's of $\underset{\sim}{\Sigma}$ differ from those of $\underset{\sim}{\rho}$. the variables are often standardized when they have different units or widely different scales.

4.5-Sample principal components

$\underset{\sim}{S} = \hat{\underset{\sim}{\Sigma}}$ $\quad$ $\hat{\underset{\sim}{\mu}} = \bar{\underset{\sim}{x}}$

Used to summarize the sample variation by PCs.

The Algebra is the same as in population principal components.

- $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are n independent observations from a population with $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.
- $\bar{\mathbf{x}}_{p\times 1}$ = sample mean vector.
- $\mathbf{S}_{p\times p} = \{s_{ik}\}$ = sample covariance matrix.
- $\mathbf{S}$ has eigenvalue/vector pairs $(\hat{\lambda}_1, \hat{\boldsymbol{\gamma}}_1), \ldots, (\hat{\lambda}_p, \hat{\boldsymbol{\gamma}}_p)$ where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_p$.
- The $\hat{}$ indicates these are estimates of population values.
- The i^{th} sample principal component is given by

$$\hat{\mathbf{y}}_i = \hat{\boldsymbol{\gamma}}_i^T \mathbf{x} = \hat{\boldsymbol{\gamma}}_{i1} x_1 + \hat{\boldsymbol{\gamma}}_{i2} x_2 + \cdots + \hat{\boldsymbol{\gamma}}_{ip} x_p$$

- The i^{th} PC sample variance = $\widehat{\text{var}}(\hat{y}_i) = \hat{\lambda}_i$ for $i = 1, \ldots, p$.
- The PC sample covariances = $\widehat{\text{cov}}(\hat{y}_i, \hat{y}_k) = 0$ for all $i \neq k$.

Total sample variance $= \sum_{i=1}^{p} s_{ii} = \widehat{\lambda}_1 + \widehat{\lambda}_2 + \cdots + \widehat{\lambda}_p$

$$r_{\widehat{y}_i, x_k} = \frac{\widehat{\boldsymbol{\gamma}}_{ik}\sqrt{\widehat{\lambda}_i}}{\sqrt{s_{kk}}}, \quad i, k = 1, 2, \ldots, p$$

$s_{ii} = s_i^2$

$\hat{\rho}_{\hat{y}_i, x_k} = r_{\hat{y}_i, x_k}$

Standardized sample principal components

Let $\mathbf{z}_j = D^{-1/2}(\mathbf{x}_j - \overline{\mathbf{x}})$, $j = 1, 2, \ldots, p$, be the standardized observations with covariance matrix $R = (D^{1/2})^{-1}S(D^{1/2})^{-1}$. Then, the ith sample principal component is given by

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

where the $(\widehat{\lambda}_i, \underset{\sim}{\widehat{\boldsymbol{\gamma}}}_i)$ are the eigenvalue-eigenvector pairs of R with $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \cdots \geq \widehat{\lambda}_p \geq 0$. Moreover,

$$\text{Sample variance}(\widehat{y}_i) = \widehat{\lambda}_i, \quad i = 1, 2, \ldots, p$$

$$\text{Sample covariance}(\widehat{y}_i, \widehat{y}_k) = 0, \quad i \neq k$$

$$\text{Total standardized sample variance} = \mathrm{tr}(R) = p = \sum_{i=1}^{p} \widehat{\lambda}_i$$

$$r_{\widehat{y}_i, z_k} = \widehat{\boldsymbol{\gamma}}_{ik}\sqrt{\widehat{\lambda}_i}$$

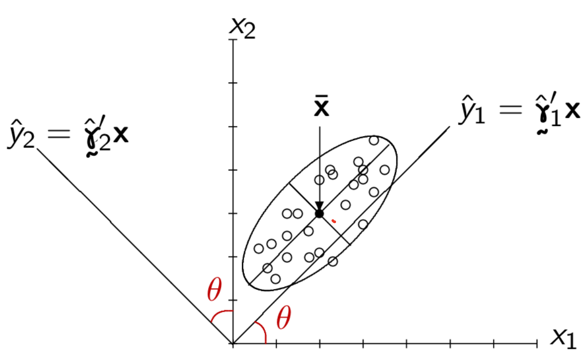
- the sample P.C based on $\underset{\sim}{S}$ are not the same of those based on $\underset{\sim}{R}$
- the variable are often standardized when they have different units and variances.

Interpretation of sample principal components

- PCs based on a sample of n p-dimensional observations are new variables specified by a rigid rotation of the original axes to a new orientation such that the directions of the axes in the new orientation have maximum variances in the sample.
  - The rotation must be rigid since the new variables must be $\perp$.
  - Directions of the new axes are based on **S** (or **R**)

The centered sample principal components $\widehat{y}_i = \widehat{\underset{\sim}{\boldsymbol{\gamma}}}_i'(\mathbf{x} - \overline{\mathbf{x}})$, $i = 1, 2, \ldots, p$, can be viewed as the result of translating the origin of the original coordinate system to $\overline{\mathbf{x}}$ and then rotating the coordinate axes until they pass through the scatter in the directions of maximum variance.

Geometry of Sample PC



The PCs are projections of observations onto the principal axes of the ellipsoids.

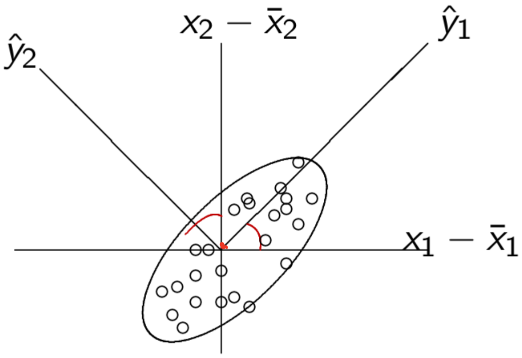
We can re-center the x's, which also centers the $\hat{y}$'s; that is

$$(\mathbf{x}_i - \bar{\mathbf{x}}) = 0 \longrightarrow \hat{y}_i \quad \text{has mean } 0$$

Subtraction of $\bar{\mathbf{x}}$ only effects the mean and does not effect variances and covariance.

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \underset{\text{shift location}}{\longrightarrow} \begin{pmatrix} x_1 - \bar{x}_1 \\ x_2 - \bar{x}_2 \end{pmatrix} \underset{\text{rigid rotatation}}{\longrightarrow} \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \end{pmatrix}$$

The PCs are projections of observations onto the principal axes of the ellipsoids.



How Many Components to Retain ??

The number of principal components

How many principal components should be retained? (No definite answer)

- The amount of total sample variance explained;
- The relative sizes of the eigenvalues;
- The subject-matter interpretations of the components

✓ A useful visual aid to determining an appropriate number of principal components is a *scree plot* (the magnitude of an eigenvalue vs. its number).

Remark: A component associated with an eigenvalue near zero and, hence, deemed unimportant, may indicate an unsuspected linear dependency in the data.

A scree plot

The number of components is taken to be the point at which the remaining eigenvalues are relatively small and all about the same size.

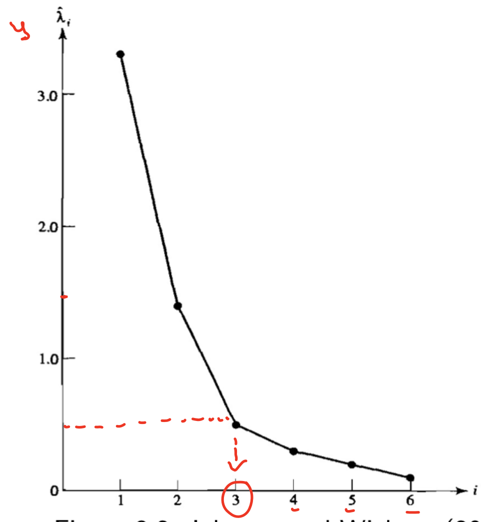


Figure 8.2, Johnson and Wichern (2007)

Some possible rules for choosing the number of P.C's :

- retain the first k PC's to explain ≈ 80% - 90% of the total variance;
- retain PC's with eigenvalues $\geq \bar{\hat{\lambda}} = \dfrac{\sum_{i=1}^{p} \hat{\lambda}_i}{p}$

using the standardized variables $\bar{\hat{\lambda}} = 1$

Kaiser rule : Keep components with $\bar{\hat{\lambda}} \geq 1$

Graphing Principal Components

• Plots of the principal components can reveal suspect observations, as well as provide checks on the assumption of normality.

- Reveal suspect observations (outliers, influential observations).
- Check multivariate normality assumptions.
- Look for clusters.
- Provide insight into structure in the data.

Suspect Observations

- The first PCs can help reveal influential observations: those that contribute more to variances than other observations such that if we removed them the results change quite a bit.
- The last PCs can help to reveal outliers: those observations that are a typical of the data set; they're inconsistent with the rest of the data (could be miss-coded).

Graphing the principal components

• To help check the normal assumption, construct scatter diagrams for pairs of the first few principal components. Also, make Q-Q plots from the sample values generated by *each* principal components.

• Construct scatter diagrams and Q-Q plots for the last few principal components. These help identify suspect observations.

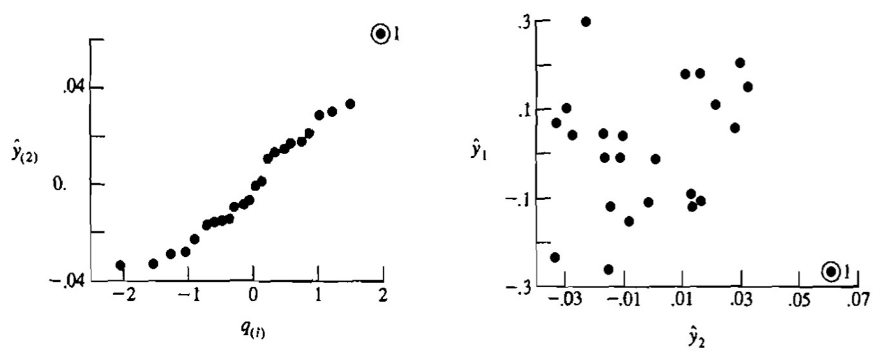


Figure 8.5 and 8.6, Johnson and Wichern (2007)

PCA as a Preliminary to Other Analysis

PCA is often used in conjunction with other data and statistical procedures, including

- Multiple regression to overcome problems of multicollinearity (use PCs as independent/predictor variables) or to select a sub-set of the original variables.
- MANOVA
- Discriminant analysis: get a lower-dimensional "look" at structure in data.
- Cluster analysis: Scaling (i.e., PCA) and clustering are often both used when concern is with finding groups of similar objects in a space.

P.C'S Explaination:

- Based on Loadings ($\underset{\sim}{\gamma}_i \Rightarrow$ weights of variables x's)
- Based on correlations between PC's and the variables (x's or z's)

Example 4. The weekly rates of return for five stocks listed on the New York Stock Exchange were determined for the period January 1975 through December 1976. Let $x_1, x_2, \ldots, x_5$ denote observed weekly rates of return for the five stocks. Then

$$\bar{x} = [0.0054, 0.0048, 0.0057, 0.0063, 0.0037]^{\tau}$$

and

$$R = \begin{bmatrix} 1.000 & 0.577 & 0.509 & 0.387 & 0.462 \\ 0.577 & 1.000 & 0.599 & 0.389 & 0.322 \\ 0.509 & 0.599 & 1.000 & 0.436 & 0.426 \\ 0.387 & 0.389 & 0.436 & 1.000 & 0.523 \\ 0.462 & 0.322 & 0.426 & 0.523 & 1.000 \end{bmatrix}$$

Different Notation:

The eigenvalues and corresponding normalized eigenvectors of R are:

$$\hat{\underset{\sim}{e}}_i = \hat{\underset{\sim}{\gamma}}_i$$

$$\hat{\lambda}_1 = 2.857 \qquad \hat{e}_1 = [0.464, 0.457, 0.470, 0.421, 0.421]^{\tau}$$

$$\hat{\lambda}_2 = 0.809 \qquad \hat{e}_2 = [0.240, 0.509, 0.260, -0.526, -0.582]^{\tau}$$

$$\hat{\lambda}_3 = 0.540 \qquad \hat{e}_3 = [-0.612, 0.178, 0.335, 0.541, -0.435]^{\tau}$$

$$\hat{\lambda}_4 = 0.452 \qquad \hat{e}_4 = [0.387, 0.206, -0.662, 0.472, -0.382]^{\tau}$$

$$\hat{\lambda}_5 = 0.343 \qquad \hat{e}_5 = [-0.451, 0.676, -0.400, -0.176, 0.385]^{\tau}$$

Using the standardized variables, we obtain the first two sample principal components

$$\hat{y}_1 = \hat{e}_1^{\tau}\boldsymbol{z} = 0.464z_1 + 0.457z_2 + 0.470z_3 + 0.421z_4 + 0.421z_5$$

$$\hat{y}_2 = \hat{e}_2^{\tau}\boldsymbol{z} = 0.240z_1 + 0.509z_2 + 0.260z_3 - 0.526z_4 - 0.582z_5$$

These components account for

$$\left(\frac{\hat{\lambda}_1 + \hat{\lambda}_2}{p}\right) \times 100\% = 73\%$$

of the total sample variance. The first component is an equally weighted sum, or "index", of the five stocks. This component might be called a market component. The second component represents a contrast between the first three stocks (which were chemical stocks) and the last two stocks (oil stocks). It might be called an industry component.

Example 5: % of People employed is different Industries

Dataset: $\begin{cases} 26 \text{ countries} \\ 3 \text{ variables} \end{cases}$

X_1 = 'percent in manufacturing'

X_2 = ' " in services Industry'

X_3 = ' " in social and personal services'

$$\underset{\sim}{S} = \begin{bmatrix} 49.109 & 6.535 & 7.379 \\ & 20.933 & 17.876 \\ & & 46.643 \end{bmatrix};$$ with eigenvectors and eigenvalues:

$\hat{\underset{\sim}{\gamma}}_1^T = (0.580; 0.396; 0.712)$; $\hat{\lambda}_1 = 62.62$

$\hat{\underset{\sim}{\gamma}}_2^T = (0.811; -0.207; -0.546)$; $\hat{\lambda}_2 = 42.47$

$\hat{\underset{\sim}{\gamma}}_3^T = (-0.069; 0.894; -0.442)$; $\hat{\lambda}_3 = 11.60$

P.C's:

$\hat{Y}_1 = \underset{\hat{\gamma}_{11}}{0.580}(X_1 - \bar{x}_1) + 0.396(X_2 - \bar{x}_2) + 0.712(X_3 - \bar{x}_3)$

$\hat{Y}_2 = 0.811(X_1 - \bar{x}_1) - 0.207(X_2 - \bar{x}_2) - 0.546(X_3 - \bar{x}_3)$

$\hat{Y}_3 = -0.069(X_1 - \bar{x}_1) + 0.894(X_2 - \bar{x}_2) - 0.442(X_3 - \bar{x}_3)$

$var(\hat{Y}_1) = \hat{\lambda}_1$

variance explained:

i	$\widehat{var}(\hat{Y}_i) = \hat{\lambda}_i$	% cumulative variance explained $= \dfrac{\sum_{j=1}^{i} \lambda_j}{\sum_{i=1}^{3} \lambda_i} \times 100\%$
1	62.62	$\dfrac{\hat{\lambda}_1}{\hat{\lambda}_1+\hat{\lambda}_2+\hat{\lambda}_3} \times 100\% = \dfrac{62.62}{116.68} \times 100\% = 53.66$
2	42.47	90.06
3	11.60	100.00
total	116.68	

$$tr(\underset{\sim}{S}) = 49.109 + 20.933 + 46.643$$

Correlations: $r_{\hat{Y}_i, X_k} = \dfrac{\hat{\gamma}_{ik}\sqrt{\hat{\lambda}_i}}{s_k}$, $i, k = 1, \dots, 3$

PC's

original variables	$\hat{Y}_1$	$\hat{Y}_2$	$\hat{Y}_3$
X_1	$\dfrac{\hat{\gamma}_{11}\sqrt{\hat{\lambda}_1}}{s_1} = \dfrac{\sqrt{62.62 \times 0.580}}{\sqrt{49.109}} = 0.66$	0.75	
X_2	$\dfrac{0.396\sqrt{62.62}}{\sqrt{20.933}} = 0.69$	-0.30	
X_3	0.82	-0.52	

Possible Interpretation of P.C's:

- $\hat{Y}_1$: all variables are contributing to the 1st P.C., it cloud be an overall percent employment in all Industries (global measure)
- $\hat{Y}_2$: Is a contrast between manufacturing (X_1) with service (X_2) and social (X_3)
- $\hat{Y}_3$: Contrast between Industry (X_2) and Social (X_3)