Principal Components

Shyh-Kang Jeng
Department of Electrical Engineering/
Graduate Institute of Communication/
Graduate Institute of Networking and Multimedia

1

Outline

- Introduction
- Popular Principal Components
- Summarizing Sample Variation by Principal Components
- Graphing the Principal Components
- Large Sample Inferences
- Monitoring Quality with Principal Components

2

Outline

- Introduction
- Popular Principal Components
- Summarizing Sample Variation by Principal Components
- Graphing the Principal Components
- Large Sample Inferences
- Monitoring Quality with Principal Components

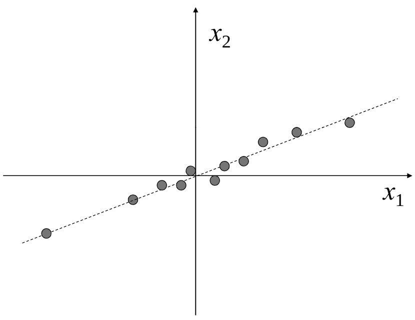
3

Questions

- What is the concept of the Principal Components?
- What are the objectives of the Principal Components?

4

Concept of Principal Components



5

Principal Component Analysis

- Explain the variance-covariance structure of a set of variables through a few linear combinations of these variables
- Objectives
  – Data reduction
  – Interpretation
- Does not need normality assumption in general

6

Outline

- Introduction
- Popular Principal Components
- Summarizing Sample Variation by Principal Components
- Graphing the Principal Components
- Large Sample Inferences
- Monitoring Quality with Principal Components

7

Questions

- How to find the Principal Components for a Random vector with a known probability distribution? (Result 8.1)
- What is the relationship between the sum of all eigenvalues and the trace of the covariance matrix? (Result 8.2)
- How to calculate the proportion of total population variance due to the kth principal component?

8

Questions

- What is the relationship between the ith principal component and the kth variable? (Result 8.3)
- What is the geometric interpretation of the principal components?
- How to find the principal components for a standardized random vector? (Result 8.4)

9

Questions

- What are the principal components for a diagonal covariance matrix?
- What are the principal components for the special covariance matrix

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \cdots & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \cdots & \rho\sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \rho\sigma^2 & \rho\sigma^2 & \cdots & \sigma^2 \end{bmatrix}$$

10

Principal Components

Random vector $\mathbf{X}' = [X_1 \quad X_2 \quad \cdots \quad X_p]$ has the covariance matrix $\mathbf{\Sigma}$

Linear combination: $Y_i = \mathbf{a}_i'\mathbf{X}, \quad i = 1, 2, \cdots, p$

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

First principal component:

$\mathbf{a}_i'\mathbf{X}$ that maximizes $\mathrm{Var}(\mathbf{a}_i'\mathbf{X})$ subject to $\mathbf{a}_i'\mathbf{a}_i = 1$

ith principal component:

$\mathbf{a}_i'\mathbf{X}$ that maximizes $\mathrm{Var}(\mathbf{a}_i'\mathbf{X})$ subject to $\mathbf{a}_i'\mathbf{a}_i = 1$ and $\mathrm{Cov}(\mathbf{a}_i'\mathbf{X}, \mathbf{a}_k'\mathbf{X}) = 0$ for $k < i$

11

Result 8.1

Covariance matrix $\mathbf{\Sigma}$ of random vector $\mathbf{X}$ is with eigenvalue-eigenvector pairs $(\lambda_i, \mathbf{e}_i)$, where $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_p \ge 0$

The ith principal component is given by

$Y_i = \mathbf{e}_i'\mathbf{X}, \quad i = 1, 2, \cdots, p, \quad$ with

$\mathrm{Var}(Y_i) = \mathbf{e}_i'\mathbf{\Sigma}\mathbf{e}_i = \lambda_i, \quad i = 1, 2, \cdots, p$

$\mathrm{Cov}(Y_i, Y_k) = \mathbf{e}_i'\mathbf{\Sigma}\mathbf{e}_k = 0, i \ne k$

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

12

Proof of Result 8.1

$\max_{\mathbf{a}\neq 0} \frac{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}}{\mathbf{a}'\mathbf{a}} = \lambda_1$ attained when $\mathbf{a} = \mathbf{e}_1$

$\mathbf{e}_1'\mathbf{e}_1 = 1$, thus $\max_{\mathbf{a}\neq 0} \frac{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}}{\mathbf{a}'\mathbf{a}} = \lambda_1 = \mathbf{e}_1'\boldsymbol{\Sigma}\mathbf{e}_1 = \mathrm{Var}(Y_1)$

$\max_{\mathbf{a}\perp \mathbf{e}_1,\ldots,\mathbf{e}_k} \frac{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}}{\mathbf{a}'\mathbf{a}} = \lambda_{k+1},\ k = 1, 2, \cdots, p-1$

$\mathbf{a} = \mathbf{e}_{k+1},\ \mathbf{e}_{k+1}'\boldsymbol{\Sigma}\mathbf{e}_{k+1} = \lambda_{k+1} = \mathrm{Var}(Y_{k+1})$

$\mathrm{Cov}(Y_i, Y_k) = \mathbf{e}_i'\boldsymbol{\Sigma}\mathbf{e}_k = \mathbf{e}_i'\lambda_k\mathbf{e}_k = 0$ for any $i \neq k$

13

Result 8.2

Covariance matrix $\boldsymbol{\Sigma}$ of random vector $\mathbf{X} = [X_1 \quad X_2 \quad \cdots \quad X_p]$ is with eigenvalue-eigenvector pairs $(\lambda_i, \mathbf{e}_i)$, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$

The ith principal component is given by

$Y_i = \mathbf{e}_i'\mathbf{X}, \quad i = 1, 2, \cdots, p,$ then

$$\sigma_{11} + \sigma_{22} + \cdots + \sigma_{pp} = \sum_{i=1}^{p} \mathrm{Var}(X_i)$$

$$= \lambda_1 + \lambda_2 + \cdots + \lambda_p = \sum_{i=1}^{p} \mathrm{Var}(Y_i)$$

14

Proof of Result 8.2

$\boldsymbol{\Sigma} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}', \quad \boldsymbol{\Lambda} = diag\{\lambda_1, \lambda_2, \cdots, \lambda_p\}$

$\mathbf{P} = [\mathbf{e}_1 \quad \mathbf{e}_2 \quad \cdots \quad \mathbf{e}_p], \quad \mathbf{P}\mathbf{P}' = \mathbf{P}'\mathbf{P} = \mathbf{I}$

$$\sigma_{11} + \sigma_{22} + \cdots + \sigma_{pp} = \sum_{i=1}^{p} \mathrm{Var}(X_i) = \mathrm{tr}(\boldsymbol{\Sigma})$$

$$= \mathrm{tr}(\mathbf{P}\boldsymbol{\Lambda}\mathbf{P}') = \mathrm{tr}(\boldsymbol{\Lambda}\mathbf{P}'\mathbf{P}) = \mathrm{tr}(\boldsymbol{\Lambda})$$

$$= \lambda_1 + \lambda_2 + \cdots + \lambda_p = \sum_{i=1}^{p} \mathrm{Var}(Y_i)$$

15

Proportion of Total Variance due to the *k*th Principal Component

$$\begin{pmatrix} \text{Proportion of total} \\ \text{population variance} \\ \text{due to the } k\text{th principal} \\ \text{component} \end{pmatrix} = \frac{\lambda_k}{\lambda_1 + \lambda_2 + \cdots + \lambda_p}$$

16

Result 8.3

$Y_i = \mathbf{e}_i'\mathbf{X}$ are the principal components obtained from the covariance matrix $\boldsymbol{\Sigma}$, then

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

are the correlation coefficients between Y_i and variable X_k. Here $\mathbf{e}_i' = [e_{i1} \quad e_{i2} \quad \cdots \quad e_{ip}]$ is the eigenvector of $\boldsymbol{\Sigma}$ corresponding to the eigenvalue λ_i. Also, $\mathbf{X}' = [X_1 \quad X_2 \quad \cdots \quad X_p]$

17

Proof of Result 8.3

$\mathbf{a}_k' = [0 \quad \cdots \quad 0 \quad 1 \quad 0 \quad \cdots \quad 0]$ so that $X_k = \mathbf{a}_k'\mathbf{X}$

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

$\mathrm{Var}(Y_i) = \lambda_i, \quad \mathrm{Var}(X_k) = \sigma_{kk}$

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

$$= \frac{\sqrt{\lambda_i}e_{ik}}{\sqrt{\sigma_{kk}}} \quad i, k = 1, 2, \cdots, p$$

18

Example 8.1

$\mathbf{X}' = [X_1 \quad X_2 \quad X_3]$ has the covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \quad \text{whose eigenvalue-eigenvector}$$

pairs are

$\lambda_1 = 5.83, \quad \mathbf{e}_1' = [0.383 \quad -0.924 \quad 0]$

$\lambda_2 = 2.00, \quad \mathbf{e}_2' = [0 \quad 0 \quad 1]$

$\lambda_1 = 0.17, \quad \mathbf{e}_1' = [0.924 \quad 0.383 \quad 0]$

19

Example 8.1

Principal components

$Y_1 = \mathbf{e}_1'\mathbf{X} = 0.383X_1 - 0.924X_2$

$Y_2 = \mathbf{e}_2'\mathbf{X} = X_3$

$Y_3 = \mathbf{e}_3'\mathbf{X} = 0.924X_1 + 0.383X_2$

Verification

$$\begin{aligned}\mathrm{Var}(Y_1) &= (0.383)^2\,\mathrm{Var}(X_1) \\ &\quad + 2(0.383)(-0.924)\,\mathrm{Cov}(X_1, X_2) + (-0.924)^2\,\mathrm{Var}(X_2) \\ &= 5.83 = \lambda_1\end{aligned}$$

$\mathrm{Cov}(Y_1, Y_2) = 0.383\,\mathrm{Cov}(X_1, X_3) - 0.924\,\mathrm{Cov}(X_2, X_3) = 0$

20

Example 8.1

$\sigma_{11} + \sigma_{22} + \sigma_{33} = 8 = 5.83 + 2.00 + 0.17 = \lambda_1 + \lambda_2 + \lambda_3$

$$\frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3} = 0.73, \quad \frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 + \lambda_3} = 0.98$$

$$\rho_{Y_1, X_1} = \frac{e_{11}\sqrt{\lambda_1}}{\sqrt{\sigma_{11}}} = \frac{0.383\sqrt{0.583}}{\sqrt{1}} = 0.925$$

$$\rho_{Y_1, X_2} = \frac{e_{12}\sqrt{\lambda_1}}{\sqrt{\sigma_{22}}} = \frac{-0.924\sqrt{0.583}}{\sqrt{5}} = -0.998$$

$$\rho_{Y_2, X_1} = \rho_{Y_2, X_2} = 0, \quad \rho_{Y_2, X_3} = \frac{\sqrt{\lambda_2}}{\sqrt{\sigma_{33}}} = 1$$

21

Geometrical Interpretation

$\mathbf{X} : N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$\boldsymbol{\Sigma}$ is with eigenvalue-eigenvector pairs $(\lambda_1, \mathbf{e}_i)$

constant probability density ellipsoid

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

$$c^2 = \frac{1}{\lambda_1}\left(\mathbf{e}_1'(\mathbf{x} - \boldsymbol{\mu})\right)^2 + \frac{1}{\lambda_2}\left(\mathbf{e}_2'(\mathbf{x} - \boldsymbol{\mu})\right)^2 + \cdots + \frac{1}{\lambda_p}\left(\mathbf{e}_p'(\mathbf{x} - \boldsymbol{\mu})\right)^2$$

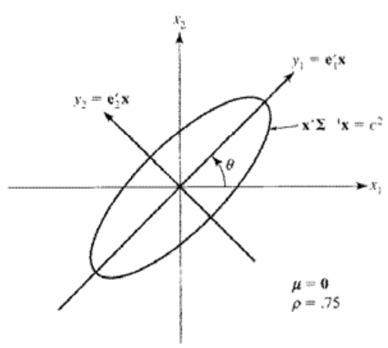
Principal components of $\mathbf{x} - \boldsymbol{\mu}$: $y_i = \mathbf{e}_i'(\mathbf{x} - \boldsymbol{\mu})$

$i = 1, 2, \cdots, p$

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

22

Geometric Interpretation



23

Standardized Variables

$$Z_i = \frac{X_i - \mu_i}{\sqrt{\sigma_{ii}}}, \quad i = 1, 2, \cdots, p$$

$$\mathbf{Z} = \mathbf{V}^{-1/2}(\mathbf{X} - \boldsymbol{\mu}), \; \mathbf{V}^{1/2} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{\sigma_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\sigma_{pp}} \end{bmatrix}$$

$$\mathrm{Cov}(\mathbf{Z}) = \mathbf{V}^{-1/2}\boldsymbol{\Sigma}\mathbf{V}^{1/2} = \boldsymbol{\rho} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{12} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1p} & \rho_{2p} & \cdots & 1 \end{bmatrix}$$

24

Result 8.4

$\mathbf{Z}' = [Z_1 \quad Z_2 \quad \cdots \quad Z_p]$ with $\text{Cov}(\mathbf{Z}) = \boldsymbol{\rho}$

$(\lambda_i, \mathbf{e}_i)$: eigenvalue-eigenvector pairs of $\boldsymbol{\rho}$

$\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_p \ge 0$

The ith principal component of $\mathbf{Z}$:

$$Y_i = \mathbf{e}_i'\mathbf{Z} = \mathbf{e}_i'\mathbf{V}^{-1/2}(\mathbf{X} - \boldsymbol{\mu}), \quad i = 1, 2, \cdots, p$$

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

$$\rho_{Y_i, Z_k} = e_{ik}\sqrt{\lambda_i}, \quad i, k = 1, 2, \cdots, p$$

25

Proportion of Total Variance due to the *k*th Principal Component

$$\begin{pmatrix} \text{Proportion of (standardized)} \\ \text{population variance} \\ \text{due to the } k\text{th principal} \\ \text{component} \end{pmatrix} = \frac{\lambda_k}{p}, \quad k = 1, 2, \cdots, p$$

26

Example 8.2

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

Eigenvalue-eigenvector pairs for $\boldsymbol{\Sigma}$:

$\lambda_1 = 100.16, \quad \mathbf{e}_1' = [0.040 \quad 0.999]$

$\lambda_2 = 0.84, \quad \mathbf{e}_2' = [0.999 \quad -0.040]$

Eigenvalue-eigenvector pairs for $\boldsymbol{\rho}$:

$\lambda_1 = 1 + \rho = 1.4, \quad \mathbf{e}_1' = [0.707 \quad 0.707]$

$\lambda_2 = 1 - \rho = 0.6, \quad \mathbf{e}_2' = [0.707 \quad -0.707]$

27

Example 8.2

Principal components for $\boldsymbol{\Sigma}$: $\dfrac{\lambda_1}{\lambda_1 + \lambda_2} = 0.992$

$Y_1 = 0.040X_1 + 0.999X_2$

$Y_2 = 0.999X_1 - 0.040X_2$

Principal components for $\boldsymbol{\rho}$: $\dfrac{\lambda_1}{p} = 0.7$

$Y_1 = 0.707Z_1 + 0.707Z_2 = 0.707(X_1 - \mu_1) + 0.0707(X_2 - \mu_2)$

$Y_2 = 0.707Z_1 - 0.707Z_2 = 0.707(X_1 - \mu_1) - 0.0707(X_2 - \mu_2)$

$\rho_{Y_1, Z_1} = e_{11}\sqrt{\lambda_1} = 0.837, \quad \rho_{Y_1, Z_2} = e_{12}\sqrt{\lambda_1} = 0.837$

28

Principal Components for Diagonal Covariance Matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & 0 & \cdots & 0 \\ 0 & \sigma_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{pp} \end{bmatrix}, \mathbf{e}_i' = [0 \ \cdots \ 0 \ 1 \ 0 \ \cdots \ 0]$$

$\boldsymbol{\Sigma}\mathbf{e}_i = \sigma_{ii}\mathbf{e}_i, \quad Y_i = \mathbf{e}_i'\mathbf{X} = X_i$

$\boldsymbol{\rho} = \mathbf{I}, \quad \boldsymbol{\rho}\mathbf{e}_i = 1\mathbf{e}_i, \quad Y_i = \mathbf{e}_i'\mathbf{Z} = Z_i$

$\mathbf{X}: N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, constant density ellipsoid is
a right ellipsoid for $\mathbf{X}$
and a sphere for $\mathbf{Z}$

29

Principal Components for a Special Covariance Matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \cdots & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \cdots & \rho\sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \rho\sigma^2 & \rho\sigma^2 & \cdots & \sigma^2 \end{bmatrix}, \boldsymbol{\rho} = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}$$

$$\lambda_1 = 1 + (p-1)\rho, \mathbf{e}_1' = \begin{bmatrix} \frac{1}{\sqrt{p}} & \frac{1}{\sqrt{p}} & \cdots & \frac{1}{\sqrt{p}} \end{bmatrix}$$

$$\lambda_2 = \lambda_3 = \cdots = \lambda_p = 1 - \rho$$

30

Principal Components for a Special Covariance Matrix

$$\mathbf{e}_i' = \left[\frac{1}{\sqrt{(i-1)i}} \quad \cdots \quad \frac{1}{\sqrt{(i-1)i}} \quad \frac{-(i-1)}{\sqrt{(i-1)i}} \quad 0 \quad \cdots \quad 0 \right]$$

$i = 2, \cdots, p$

$$Y_1 = \mathbf{e}_1'\mathbf{Z} = \frac{1}{\sqrt{p}} \sum_{i=1}^{p} Z_i, \qquad \frac{\lambda_1}{p} = \rho + \frac{1-\rho}{p}$$

the last $p-1$ components collectively contribute very little to the total variance and can be neglected when ρ is near 1

31

Outline

- Introduction
- Popular Principal Components
- Summarizing Sample Variation by Principal Components
- Graphing the Principal Components
- Large Sample Inferences
- Monitoring Quality with Principal Components

32

Questions

- What are the sample principal components?
- How to compute the sample principal components?
- How to decide the number of principal components required?
- What is the geometric interpretation of the sample principal components?

33

Questions

- How to compute the sample principal components for standardized random vectors?
- What does it mean for an unusually small value for the last eigenvalue from either the sample covariance or correlation matrix?

34

Sample Principal Components

$\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$: n independent drawings from some $p-$dimensional population with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$

sample mean $\bar{\mathbf{x}}$, sample covariance matrix $\mathbf{S}$

first sample principal component $\mathbf{a}_1'\mathbf{x}_j$:

$$\max_{\mathbf{a}_1} \mathbf{a}_1'\mathbf{S}\mathbf{a}_1 \quad \text{subject to } \mathbf{a}_1'\mathbf{a}_1 = 1$$

ith sample principal component $\mathbf{a}_i'\mathbf{x}_j$:

$$\max_{\mathbf{a}_i} \mathbf{a}_i'\mathbf{S}\mathbf{a}_i \quad \text{subject to } \mathbf{a}_i'\mathbf{a}_i = 1 \text{ and } \mathbf{a}_i'\mathbf{S}\mathbf{a}_k = 0$$

35

Sample Principal Components

$\mathbf{S} = \{s_{ik}\}$ is with eigenvalue-eigenvector pairs $(\hat{\lambda}_i, \hat{\mathbf{e}}_i)$, $\quad i, k = 1, 2, \cdots, p$

ith sample principal component of observation $\mathbf{x}$:

$$\hat{y}_i = \hat{\mathbf{e}}_i'\mathbf{x} = \hat{e}_{i1}x_1 + \hat{e}_{i2}x_2 + \cdots + \hat{e}_{ip}x_p$$

$$\hat{\lambda}_1 \ge \hat{\lambda}_2 \ge \cdots \ge \hat{\lambda}_p \ge 0$$

sample variance$(\hat{y}_k) = \hat{\lambda}_k$

sample covariance$(\hat{y}_i, \hat{y}_k) = 0, \quad i \ne k$

$$\text{Total sample variance} = \sum_{i=1}^{p} s_{ii} = \sum_{i=1}^{p} \hat{\lambda}_i, \quad r_{\hat{y}_i, x_k} = \frac{\hat{e}_{ik}\sqrt{\hat{\lambda}_i}}{\sqrt{s_{kk}}}$$

36

Example 8.3

Socioeconomic variables for 61 tracts in Madison, Wisconsin.

X_1 : total population (thousands)

X_2 : professional degree (percent)

X_3 : employed age over 16 (percent)

X_4 : government employment (percent)

X_5 : median home value ($10,000s)

$$\bar{\mathbf{x}}' = \begin{bmatrix} 4.47 & 3.96 & 71.42 & 26.91 & 1.64 \end{bmatrix}$$

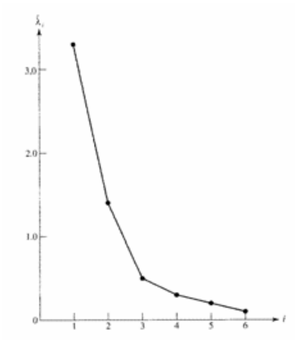
$$\mathbf{S} = \begin{bmatrix} 3.397 & & & & \\ -1.102 & 9.673 & & & \\ 4.306 & -1.513 & 55.626 & & \\ -2.078 & 10.953 & -28.937 & 89.067 & \\ 0.027 & 1.203 & -0.044 & 0.957 & 0.319 \end{bmatrix}$$

Example 8.3

Coefficients for the Principal Components
(Correlation Coefficients in Parentheses)

Variable	$\hat{\mathbf{e}}_1$ $(r_{\hat{y}_1, x_k})$	$\hat{\mathbf{e}}_2$ $(r_{\hat{y}_2, x_k})$	$\hat{\mathbf{e}}_3$	$\hat{\mathbf{e}}_4$	$\hat{\mathbf{e}}_5$
Total population	−0.039(−.22)	0.071(.24)	0.188	0.977	−0.058
Profession	0.105(.35)	0.130(.26)	−0.961	0.171	−0.139
Employment (%)	−0.492(−.68)	0.864(.73)	0.046	−0.091	0.005
Government employment (%)	0.863(.95)	0.480(.32)	0.153	−0.030	0.007
Medium home value	0.009(.16)	0.015(.17)	−0.125	0.082	0.989
Variance ($\hat{\lambda}_i$):	107.02	39.67	8.37	2.87	0.15
Cumulative percentage of total variance	67.7	92.8	98.1	99.9	1.000

Scree Plot to Determine Number of Principal Components



Example 8.4: Pained Turtles

natural logarithms of the measured carapace length, width, and weight of 24 male pained turtles

sample mean vector :

$$\bar{\mathbf{x}} = \begin{bmatrix} 4.725 & 4.478 & 3.703 \end{bmatrix}$$

sample covariance matrix

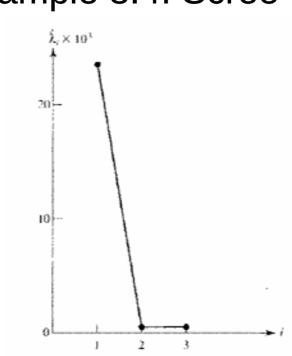
$$\mathbf{S} = 10^{-3} \begin{bmatrix} 11.072 & 8.019 & 8.160 \\ 8.019 & 6.417 & 6.005 \\ 8.160 & 6.005 & 6.773 \end{bmatrix}$$

Example 8.4

COEFFICIENTS FOR PRINCIPAL COMPONENTS
(Correlation Coefficients in Parentheses)

Variable	$\hat{\mathbf{e}}_1$ $(r_{\hat{y}_1, x_k})$	$\hat{\mathbf{e}}_2$	$\hat{\mathbf{e}}_3$
ln (length)	.683 (.99)	−.159	−.713
ln (width)	.510 (.97)	−.594	.622
ln (height)	.523 (.97)	.788	.324
Variance ($\hat{\lambda}_i$):	23.30×10^{-3}	$.60 \times 10^{-3}$	$.36 \times 10^{-3}$
Cumulative percentage of total variance	96.1	98.5	100

Example 8.4: Scree Plot



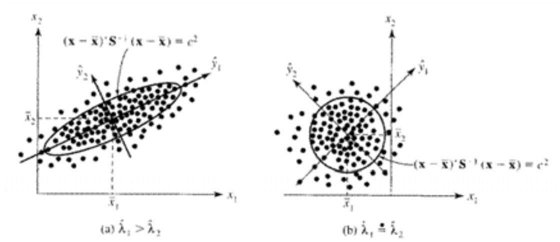
Example 8.4: Principal Component

- One dominant principal component
  - Explains 96% of the total variance
- Interpretation

$$\hat{y}_1 = 0.683\ln(length) + 0.510\ln(width) + 0.523\ln(height)$$
$$= \ln\left[(length)^{0.683}(width)^{0.510}(height)^{0.523}\right]$$
$$= \ln\left(volume \text{ of a box with adjusted dimension}\right)$$

43

Geometric Interpretation



44

Standardized Variables

$$z_{ji} = \frac{x_{ji} - \bar{x}_i}{\sqrt{s_{ii}}}, \quad i = 1, 2, \cdots, p, \quad \mathbf{Z} = \{z_{ji}\}$$

$$\mathbf{z}_j = \mathbf{D}^{-1/2}\left(\mathbf{x}_j - \bar{\mathbf{x}}\right), \quad \bar{\mathbf{z}} = \frac{1}{n}\mathbf{Z}'\mathbf{1} = 0$$

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

45

Principal Components

$\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_n$ are standardized observations with sample covariance matrix $\mathbf{R}$

$(\hat{\lambda}_i, \hat{\mathbf{e}}_i)$: eigenvalue-eigenvector pairs of $\mathbf{R}$

$\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_p \geq 0$

The ith principal component of $\mathbf{z}$:

$\hat{y}_i = \mathbf{e}_i'\mathbf{z}, \quad i = 1, 2, \cdots, p$

sample variance $(\hat{y}_i) = \hat{\lambda}_i$, sample covariance$(\hat{y}_i, \hat{y}_k) = 0, i \neq k$

total sample variance = tr$(\mathbf{R}) = p$

$r_{y_i, z_k} = \hat{e}_{ik}\sqrt{\hat{\lambda}_i}, \quad i, k = 1, 2, \cdots, p$

46

Proportion of Total Variance due to the *k*th Principal Component

$$\begin{pmatrix} \text{Proportion of (standardized)} \\ \text{sample variance} \\ \text{due to the } k\text{th sample principal} \\ \text{component} \end{pmatrix} = \frac{\hat{\lambda}_k}{p}, \quad k = 1, 2, \cdots, p$$

47

Example 8.5: Stocks Data

- Weekly rates of return for five stocks
  - X_1: JP Morgan
  - X_2: Citibank
  - X_3: Wells Fargo
  - X_4: Royal Dutch Shell
  - X_5: ExxonMobil

48

Example 8.5

$\bar{\mathbf{x}}' = [0.0011 \quad 0.0007 \quad 0.0016 \quad 0.0040 \quad 0.0040]$

$$\mathbf{R} = \begin{bmatrix} 1 & & & & \\ 0.632 & 1 & & & \\ 0.511 & 0.574 & 1 & & \\ 0.115 & 0.322 & 0.183 & 1 & \\ 0.155 & 0.213 & 0.146 & 0.683 & 1 \end{bmatrix}$$

$\hat{\lambda}_1 = 2.437, \quad \hat{\mathbf{e}}_1' = [0.469 \quad 0.532 \quad 0.465 \quad 0.387 \quad 0.361]$

$\hat{\lambda}_2 = 1.407, \quad \hat{\mathbf{e}}_2' = [-0.368 \quad -0.236 \quad -0.315 \quad 0.585 \quad 0.606]$

$\hat{\lambda}_3 = 0.501, \quad \hat{\mathbf{e}}_3' = [-0.604 \quad -0.136 \quad 0.772 \quad 0.093 \quad -0.109]$

$\hat{\lambda}_4 = 0.400, \quad \hat{\mathbf{e}}_4' = [0.363 \quad -0.629 \quad 0.289 \quad -0.381 \quad 0.493]$

$\hat{\lambda}_5 = 0.255, \quad \hat{\mathbf{e}}_5' = [0.384 \quad -0.496 \quad 0.071 \quad 0.595 \quad -0.498]$

49

Example 8.5

First two principal components:

$\hat{y}_1 = \hat{\mathbf{e}}_1'\mathbf{z} = 0.469z_1 + 0.532z_2 + 0.465z_3 + 0.387z_4 + 0.361z_5$

$\hat{y}_2 = \hat{\mathbf{e}}_2'\mathbf{z} = -0.368z_1 - 0.236z_2 - 0.315z_3 + 0.585z_4 + 0.606z_5$

$\frac{\hat{\lambda}_1 + \hat{\lambda}_2}{p} = 77\%$

$\hat{y}_1$: roughly equally weighted sum (index) of the five stocks (general stock-market component, or, market component)

$\hat{y}_2$: contrast banking stocks and the oil stocks (industry component)

50

Example 8.6

- Body weight (in grams) for n=150 female mice were obtained after the birth of their first 4 litters

$\bar{\mathbf{x}}' = [39.88 \quad 45.08 \quad 48.11 \quad 49.95]$

$$\mathbf{R} = \begin{bmatrix} 1 & & & \\ 0.7501 & 1 & & \\ 0.6329 & 0.6925 & 1 & \\ 0.6363 & 0.7386 & 0.6625 & 1 \end{bmatrix}$$

51

Example 8.6

$\hat{\lambda}_1 = 3.085, \quad \hat{\lambda}_2 = 0.382, \quad \hat{\lambda}_3 = 0.342, \quad \hat{\lambda}_4 = 0.217$

$\hat{\lambda}_1 \approx 1 + (p-1)\bar{r} = 1 + (4-1) \times 0.6854 = 3.056$

$\hat{\lambda}_2 \approx \hat{\lambda}_3 \approx \hat{\lambda}_4 << \hat{\lambda}_1$

$\hat{y}_1 = \hat{\mathbf{e}}_1'\mathbf{z} = 0.49z_1 + 0.52z_2 + 0.49z_3 + 0.50z_4$

$\frac{\hat{\lambda}_1}{p} = 0.76$

52

Comment

- An unusually small value for the last eigenvalue from either the sample covariance or correlation matrix can indicate an unnoticed linear dependency of the data set
- One or more of the variables is redundant and should be deleted
- Example: $x_4 = x_1 + x_2 + x_3$

53

Outline

- Introduction
- Popular Principal Components
- Summarizing Sample Variation by Principal Components
- Graphing the Principal Components
- Large Sample Inferences
- Monitoring Quality with Principal Components

54

Questions

- Why to check the normality of the first few principal components?
- How to pinpoint suspect observation?

55

Check Normality and Suspect Observations

- Construct scatter diagram for pairs of the first few principal components
- Make *Q-Q* plots from the sample values generated by each principal component
- Construct scatter diagram and *Q-Q* plots for the last few principal components

56

Example 8.7: Turtle Data

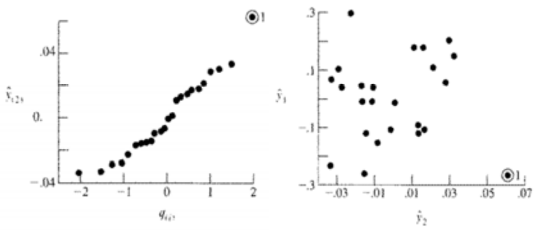
$$\hat{y}_1 = 0.683(x_1 - 4.725) + 0.510(x_2 - 4.478) + 0.523(x_3 - 3.703)$$

$$\hat{y}_2 = -0.159(x_1 - 4.725) - 0.594(x_2 - 4.478) + 0.788(x_3 - 3.703)$$

$$\hat{y}_3 = -0.713(x_1 - 4.725) + 0.622(x_2 - 4.478) + 0.324(x_3 - 3.703)$$

57

Example 8.7



58

Outline

- Introduction
- Popular Principal Components
- Summarizing Sample Variation by Principal Components
- Graphing the Principal Components
- Large Sample Inferences
- Monitoring Quality with Principal Components

59

Questions

- What are the large sample distribution for eigenvalues and eigenvectors?
- How to determine the confidence interval for an eigenvalue?
- What is the approximate distribution for estimated eigenvectors?
- How to test for equal correlation structure?

60

Large Sample Distribution for Eigenvalues and Eigenvectors

$\mathbf{S}$ is with eigen values $\hat{\boldsymbol{\lambda}}' = [\hat{\lambda}_1 \ \cdots \ \hat{\lambda}_p]$ and eigenvectors $\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \cdots, \hat{\mathbf{e}}_p$

Let $\boldsymbol{\Lambda} = diag\{\lambda_1, \cdots, \lambda_p\}$, λ_i's are eigenvalues of $\boldsymbol{\Sigma}$

$\Rightarrow \sqrt{n}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})$: approximately $N_p(\mathbf{0}, 2\boldsymbol{\Lambda}^2)$

Let $\mathbf{E}_i = \lambda_i \sum_{\substack{k=1 \\ k \neq i}}^{p} \frac{\lambda_k}{(\lambda_k - \lambda_i)^2} \mathbf{e}_k \mathbf{e}_k'$

$\Rightarrow \sqrt{n}(\hat{\mathbf{e}}_i - \mathbf{e}_i)$: approximately $N_p(\mathbf{0}, \mathbf{E}_i)$

$\hat{\lambda}_i$ is independent of the elements of associated $\hat{\mathbf{e}}_i$

61

Confidence Interval for λ_i

$\hat{\lambda}_i : N(\lambda_i, 2\lambda_i^2 / n)$ for n large

$$P\left[\frac{|\hat{\lambda}_i - \lambda_i|}{\lambda_i \sqrt{\frac{2}{n}}} \le z\left(\frac{\alpha}{2}\right)\right] = 1 - \alpha$$

$100(1-\alpha)\%$ confidence interval for λ_i :

$$\frac{\hat{\lambda}_i}{1 + z(\alpha/2)\sqrt{2/n}} \le \lambda_i \le \frac{\hat{\lambda}_i}{1 - z(\alpha/2)\sqrt{2/n}}$$

62

Approximate Distribution of Estimated Eigenvectors

$\sqrt{n}(\hat{\mathbf{e}}_i - \mathbf{e}_i)$: approximate $N_p(\mathbf{0}, \mathbf{E}_i)$

$\mathbf{E}_i$ can be approximated by

$$\hat{\mathbf{E}}_i = \hat{\lambda}_i \sum_{\substack{k=1 \\ k \neq i}}^{p} \frac{\hat{\lambda}_k}{(\hat{\lambda}_k - \hat{\lambda}_i)^2} \hat{\mathbf{e}}_k \hat{\mathbf{e}}_k'$$

$\hat{e}_{ik} : N(e_{ik}, \hat{E}_{i,kk} / n)$

63

Example 8.8

Stock price data : $N_5(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$\boldsymbol{\Sigma}$ has distinct eigenvalues $\lambda_1 > \lambda_2 > \cdots > \lambda_5 > 0$

$n = 103$ large

$\hat{\lambda}_1 = 0.0014, \quad z(0.025) = 1.96$

95% confidence interval

$$\frac{0.0014}{1 + 1.96\sqrt{2/103}} \le \lambda_1 \le \frac{0.0014}{1 - 1.96\sqrt{2/103}}, \text{ or}$$

$0.0011 \le \lambda_1 \le 0.0019$

64

Testing for Equal Correlation

$$H_0 : \boldsymbol{\rho} = \boldsymbol{\rho}_0 = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}, \quad H_1 : \boldsymbol{\rho} \neq \boldsymbol{\rho}_0$$

$$\bar{r}_k = \frac{1}{p-1} \sum_{\substack{i=1 \\ i \neq k}}^{p} r_{ik}, \quad \bar{r} = \frac{2}{p(p-2)} \sum_k \sum_{i<k} r_{ik}, \quad \hat{\gamma} = \frac{(p-1)^2 [1 - (1-\bar{r})^2]}{p - (p-2)(1-\bar{r})^2}$$

Reject H_0 in favor of H_1 if

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

65

Example 8.9

Example 8.6, female mice data $\mathbf{R} = \begin{bmatrix} 1 & & & \\ 0.7501 & 1 & & \\ 0.6329 & 0.6925 & 1 & \\ 0.6363 & 0.7386 & 0.6625 & 1 \end{bmatrix}$

$\bar{r}_1 = 0.6731, \bar{r}_2 = 0.7271, \bar{r}_3 = 0.6626, \bar{r}_4 = 0.6791, \bar{r} = 0.6855$

$\sum_k \sum_{i<k} (r_{ik} - \bar{r})^2 = 0.01277, \sum_{k=1}^{4} (\bar{r}_k - \bar{r})^2 = 0.00245, \hat{\gamma} = 2.1329$

$T = \frac{(150-1)}{(1-0.6855)^2} [0.01277 - (2.1329)(0.00245)] = 11.4$

$> \chi^2_{(4+1)(4-2)/2}(0.05) = 11.07$

The evidence against H_0 is strong, but not overwhelming

66

Outline

- Introduction
- Popular Principal Components
- Summarizing Sample Variation by Principal Components
- Graphing the Principal Components
- Large Sample Inferences
- Monitoring Quality with Principal Components

67

Questions

- How to monitor a stable process using the first two principal components?
- How to monitor a stable process using the T^2 chart from the principal components?
- How to control future values by principal components?

68

Questions

- Why avoiding Computation with Small Eigenvalues?

69

Monitoring Stable Process: Part 1

The values of the first two principal components should be stable for a process stable over time

Construct the quality ellipse for the first two principal components when n large:

$$\frac{\hat{y}_1^2}{\hat{\lambda}_1} + \frac{\hat{y}_2^2}{\hat{\lambda}_2} \le \chi_2^2(\alpha)$$

70

Example 8.10 Police Department Data

Variable	$\hat{\mathbf{e}}_1$	$\hat{\mathbf{e}}_2$	$\hat{\mathbf{e}}_3$	$\hat{\mathbf{e}}_4$	$\hat{\mathbf{e}}_5$
Appearances overtime (x_1)	.046	−.048	.629	−.643	.432
Extraordinary event (x_2)	.039	.985	−.077	−.151	−.007
Holdover hours (x_3)	−.658	.107	.582	.250	−.392
COA hours (x_4)	.734	.069	.503	.397	−.213
Meeting hours (x_5)	−.155	.107	.081	.586	.784
$\hat{\lambda}_i$	2,770,226	1,429,206	628,129	221,138	99,824

*First two sample cmponents explain 82% of the total variance

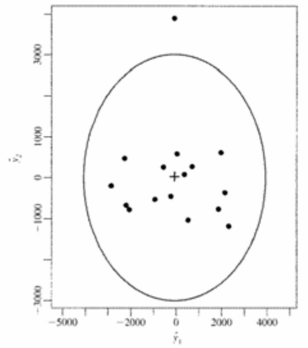
71

Example 8.10: Principal Components

Period	$\hat{y}_{j1}$	$\hat{y}_{j2}$	$\hat{y}_{j3}$	$\hat{y}_{j4}$	$\hat{y}_{j5}$
1	2044.9	588.2	425.8	−189.1	−209.8
2	−2143.7	−686.2	883.6	−565.9	−441.5
3	−177.8	−464.6	707.5	736.3	38.2
4	−2186.2	450.5	−184.0	443.7	−325.3
5	−878.6	−545.7	115.7	296.4	437.5
6	563.2	−1045.4	281.2	620.5	142.7
7	403.1	66.8	340.6	−135.5	521.2
8	−1988.9	−801.8	−1437.3	−148.8	61.6
9	132.8	563.7	125.3	68.2	611.5
10	−2787.3	−213.4	7.8	169.4	−202.3
11	283.4	3936.9	−0.9	276.2	−159.6
12	761.6	256.0	−2153.6	−418.8	28.2
13	−498.3	244.7	966.5	−1142.3	182.6
14	2366.2	−1193.7	−165.5	270.6	−344.9
15	1917.8	−782.0	−82.9	−196.8	−89.9
16	2187.7	−373.8	170.1	−84.1	−250.2

72

Example 8.10: 95% Control Ellipse



73

Monitoring Stable Process: Part 2

$$\mathbf{X} : N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad \mathbf{E} = [\mathbf{e}_1 \quad \mathbf{e}_2 \quad \cdots \quad \mathbf{e}_p]$$

$$\mathbf{X} - \boldsymbol{\mu} = \sum_{i=1}^{p} (\mathbf{X} - \boldsymbol{\mu})' \mathbf{e}_i \mathbf{e}_i = \sum_{i=1}^{p} Y_i \mathbf{e}_i$$

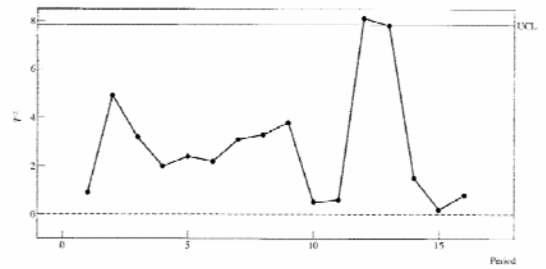
$$\mathbf{E}'(\mathbf{X} - \boldsymbol{\mu} - Y_1\mathbf{e}_1 - Y_2\mathbf{e}_2) = [0 \quad 0 \quad Y_3 \quad \cdots \quad Y_p]' = [0 \quad 0 \quad \mathbf{Y}_{(2)}]'$$

$$\mathbf{Y}_{(2)}' \boldsymbol{\Sigma}^{-1}_{\mathbf{Y}_{(2)}, \mathbf{Y}_{(2)}} \mathbf{Y}_{(2)} = \frac{Y_3^2}{\lambda_3} + \frac{Y_4^2}{\lambda_4} + \cdots + \frac{Y_p^2}{\lambda_p} : \chi^2_{p-2}$$

$$T_j^2 = \frac{\hat{y}_{j3}^2}{\hat{\lambda}_3} + \frac{\hat{y}_{j4}^2}{\hat{\lambda}_4} + \cdots + \frac{\hat{y}_{jp}^2}{\hat{\lambda}_p}, \quad \text{UCL} = \chi^2_{p-2}(\alpha)$$

74

Example 8.11 T^2 Chart for Unexplained Data



75

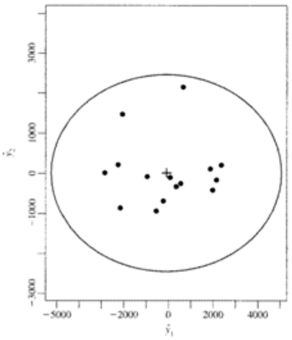
Example 8.12 Control Ellipse for Future Values

	$\hat{\mathbf{e}}_1$	$\hat{\mathbf{e}}_2$	$\hat{\mathbf{e}}_3$	$\hat{\mathbf{e}}_4$	$\hat{\mathbf{e}}_5$
Appearances overtime (x_1)	.049	.629	.304	.479	.530
Extraordinary event (x_2)	.007	−.078	.939	−.260	−.212
Holdover hours (x_3)	−.662	.582	−.089	−.158	−.437
COA hours (x_4)	.731	:503	−.123	−.336	−.291
Meeting hours (x_5)	−.159	.081	−.058	−.752	.632
$\hat{\lambda}_i$	2,964,749.9	672,995.1	396,596.5	194,401.0	92,760.3

*Example 8.10 data after dropping out-of-control case

76

Example 8.12 99% Prediction Ellipse



77

Avoiding Computation with Small Eigenvalues

$$d_{Uj}^2 = (\bar{\mathbf{x}}_j - \bar{\mathbf{x}} - \hat{y}_{j1}\hat{\mathbf{e}}_1 - \hat{y}_{j2}\hat{\mathbf{e}}_2)'(\bar{\mathbf{x}}_j - \bar{\mathbf{x}} - \hat{y}_{j1}\hat{\mathbf{e}}_1 - \hat{y}_{j2}\hat{\mathbf{e}}_2)$$

$$= (\bar{\mathbf{x}}_j - \bar{\mathbf{x}} - \hat{y}_{j1}\hat{\mathbf{e}}_1 - \hat{y}_{j2}\hat{\mathbf{e}}_2)'\hat{\mathbf{E}}\hat{\mathbf{E}}'(\bar{\mathbf{x}}_j - \bar{\mathbf{x}} - \hat{y}_{j1}\hat{\mathbf{e}}_1 - \hat{y}_{j2}\hat{\mathbf{e}}_2)$$

$$= \sum_{k=3}^{p} \hat{y}_{jk}^2 : \text{approximate } c\chi^2_\nu$$

$$\bar{d}_U^2 = \frac{1}{n}\sum_{j=1}^{n} d_U^2 = c\nu, \quad s_{d^2}^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(d_{Uj}^2 - \bar{d}_U^2\right)^2 = 2c^2\nu$$

$$c = \frac{s_{d^2}^2}{2\bar{d}_U^2}, \quad \nu = 2\frac{\left(\bar{d}_U^2\right)^2}{s_{d^2}^2}$$

78