연세대학교 통계 데이터 사이언스 학회 ESC 23-2 FALL WEEK2

Principal Component Analysis

[ESC 정규세션 학술부] 노희준 최영준

Contents

1.PCA

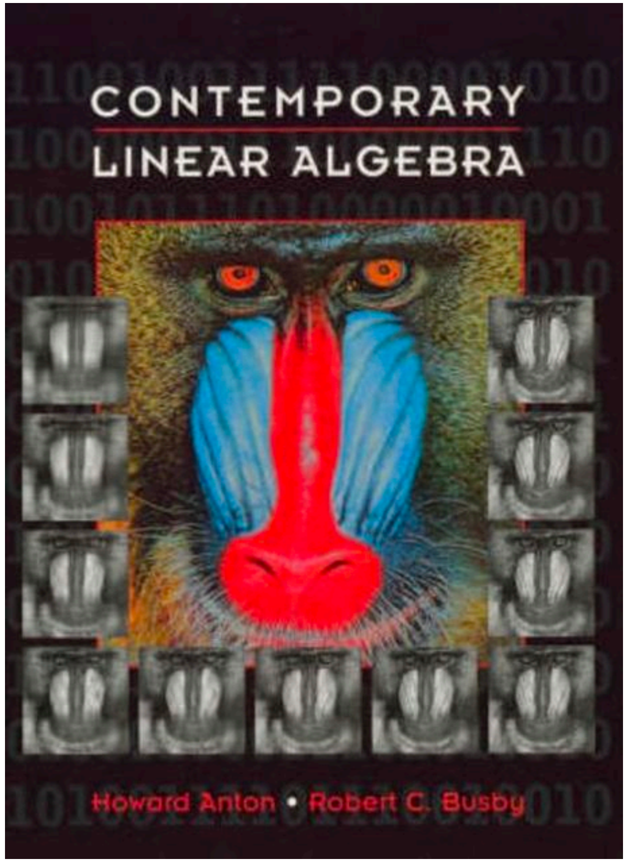
1. PCA

PCA의 목적은 data reduction

observations의 차원을 줄이는 것

1) 분석을 위해

2) 데이터의 용량을 줄이기 위해



1.PCA

How to reduce

다양한 방법으로 dimension reduction할 수 있지만 조금 더 reasonable한 방식으로 줄여야 함

-> weighted average 방식으로 $X_1, \ldots, X_p$에 각기 가중치를 부여하여 p차원의 데이터를 줄일 수 있음.

$$\delta^\top X = \sum_{j=1}^{p} \delta_j X_j \quad \text{so that} \quad \sum_{j=1}^{p} \delta_j^2 = 1.$$

Standardized linear combination

$$\max_{\{\delta:\|\delta\|=1\}} Var(\delta^\top X) = \max_{\{\delta:\|\delta\|=1\}} \delta^\top \, Var(X)\delta.$$

1. PCA

How to reduce

우리가 궁금한 것은 δ의 direction인데,

이때 δ의 방향은 X의 covariance matrix의 spectral decomposition을 통해서 찾을 수 있음

THEOREM 2.5 If $\mathcal{A}$ and $\mathcal{B}$ are symmetric and $\mathcal{B} > 0$, then the maximum of $x^\top \mathcal{A} x$ under the constraints $x^\top \mathcal{B} x = 1$ is given by the largest eigenvalue of $\mathcal{B}^{-1}\mathcal{A}$. More generally,

$$\max_{\{x : x^\top \mathcal{B} x = 1\}} x^\top \mathcal{A} x = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p = \min_{\{x : x^\top \mathcal{B} x = 1\}} x^\top \mathcal{A} x,$$

$(\mathcal{A} = Var(X) = \Sigma, \ \mathcal{B} = I, \ x = \delta)$

1. PCA

How to reduce

이를 통해 delta의 방향은 covariance matrix의 가장 큰 eigenvalue λ_1에 대응하는 eigenvector γ_1이다.

그림 4. 공분산행렬 Matrix 1의 각 원소들이 의미하는 것

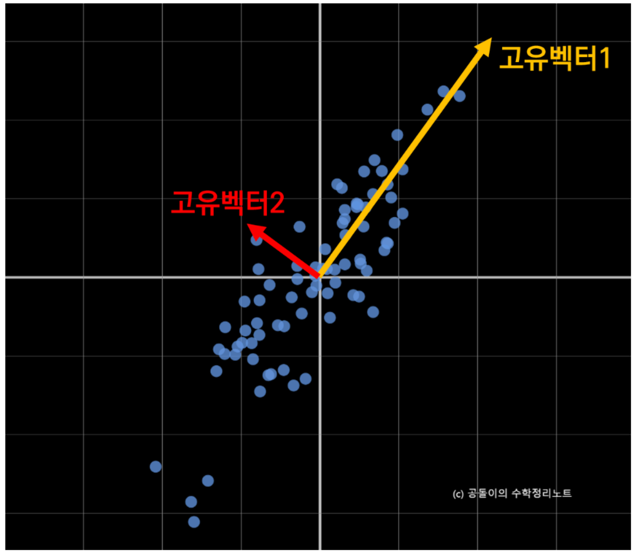
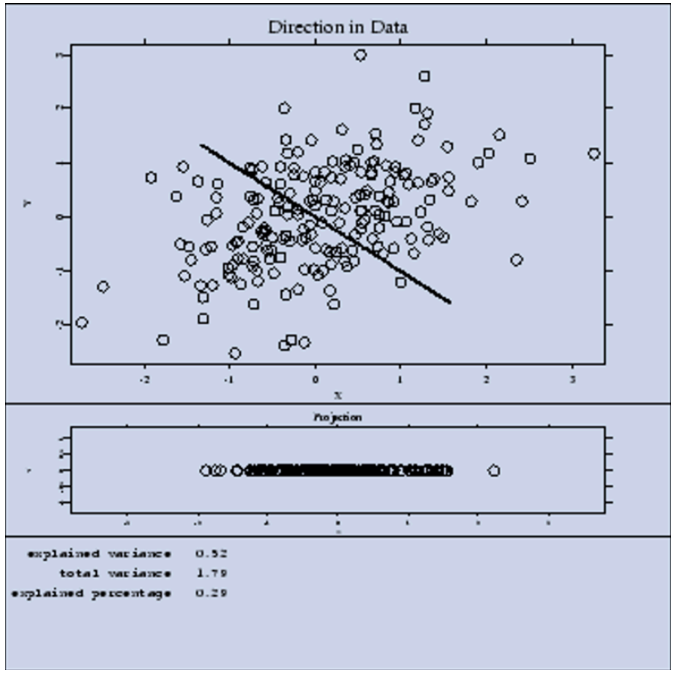


그림 5. 공분산행렬 Matrix 1의 고유벡터

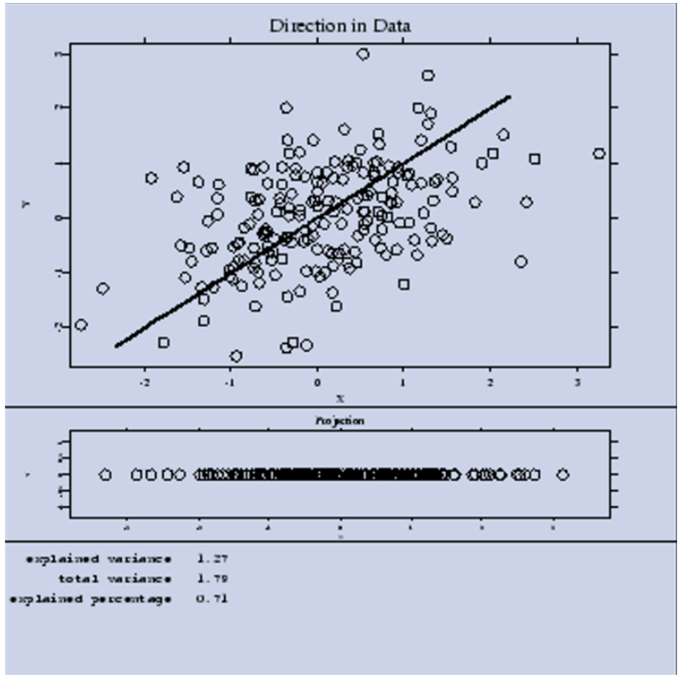
고유벡터는 행렬이 벡터에 작용하는 주축의 방향, 고윳값은 고유벡터 방향으로 얼마나 늘려져 있는지에 대한 정도

1. PCA

How to reduce



가장 큰 variance를 차지하는 first PC



1. PCA

How to reduce

이렇게 구한 eigenvector γ에 대해 X와 내적하여 PC로 나타내면

$y_1 = \gamma_1^T X, y_2 = \gamma_2^T X$

γ_1과 γ_2는 orthogonal

Matrix form으로 나타내면 $Y = \Gamma^T (X - \mu)$

이때 μ를 빼주는 이유는 mean이 0이고 centered data여야 하기 때문

1. PCA

How to reduce

Example)

다음과 같은 Bivariate normal distribution을 따르는 X가 있다고 가정

$N(0,\Sigma)$ with $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \rho > 0$

Σ의 eigenvalue를 구하면 $\begin{vmatrix} 1-\lambda & \rho \\ \rho & 1-\lambda \end{vmatrix} = 0$를 만족시키는

$\lambda_1 = 1 + \rho,\ \lambda_2 = 1 - \rho$가 X의 covariance matrix의 eigenvalue이다.

1. PCA

How to reduce

Eigenvalue λ_1에 해당하는 eigenvector를 구하면

$$\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (1+\rho)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$\begin{aligned} x_1 + \rho x_2 &= x_1 + \rho x_1 \\ \rho x_1 + x_2 &= x_2 + \rho x_2 \end{aligned}$$

$$\gamma_1 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}.$$

$$\Gamma = (\gamma_1, \gamma_2) = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}$$

γ_1과 γ_2는 orthogonal하기 때문

1. PCA

How to reduce

이렇게 구한 PC transformation을 나타내면

$$Y = \Gamma^\top (X - \mu) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} X$$

이때 PC variance는

$$\begin{aligned} Var(Y_1) &= Var\left\{\frac{1}{\sqrt{2}}(X_1 + X_2)\right\} = \frac{1}{2}\, Var(X_1 + X_2) \\ &= \frac{1}{2}\{Var(X_1) + Var(X_2) + 2\, Cov(X_1, X_2)\} \\ &= \frac{1}{2}(1 + 1 + 2\rho) = 1 + \rho \\ &= \lambda_1. \end{aligned}$$

$$Var(Y_2) = \lambda_2$$

1. PCA

How to reduce

THEOREM 9.1 For a given $X \sim (\mu, \Sigma)$ let $Y = \Gamma^\top (X - \mu)$ be the PC transformation. Then

$$EY_j = 0, \quad j = 1, \dots, p$$

$$Var(Y_j) = \lambda_j, \qquad j = 1, \dots, p$$

$$Cov(Y_i, Y_j) = 0, \qquad i \neq j$$

$$Var(Y_1) \geq Var(Y_2) \geq \cdots \geq Var(Y_p) \geq 0$$

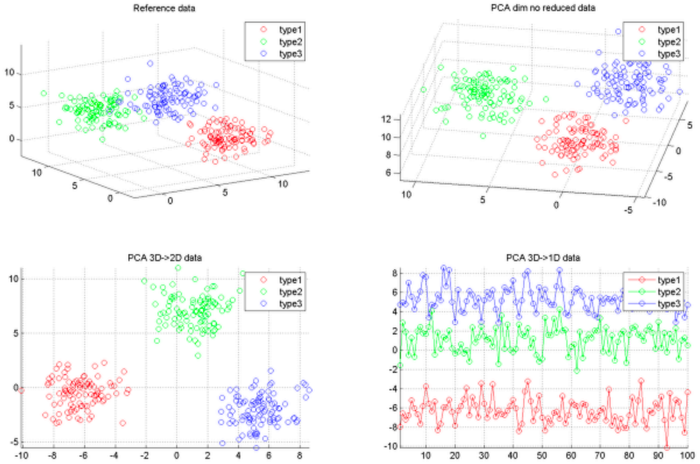
$$\sum_{j=1}^{p} Var(Y_j) = \mathrm{tr}(\Sigma)$$

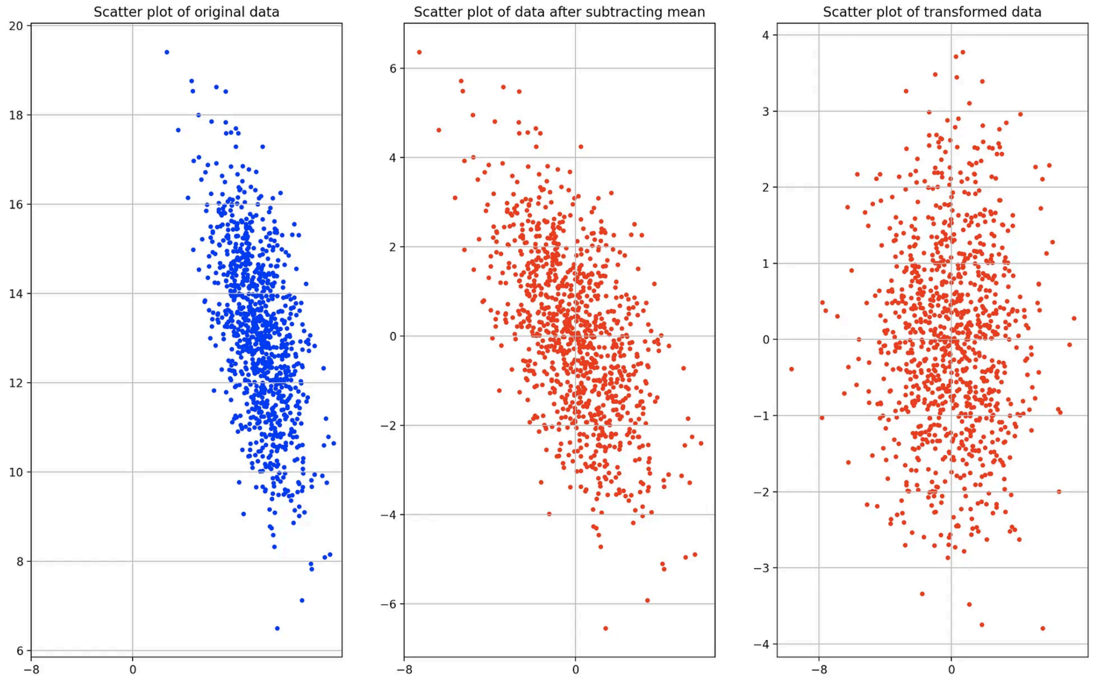
$$\prod_{j=1}^{p} Var(Y_j) = |\Sigma|.$$

1. PCA

How to reduce

그렇게 PC transformation을 하게 되면 다음과 같이 데이터를 reduction할 수 있게 됨





Change of basis의 효과(reconstruct도 가능)

1. PCA

Reconstruct

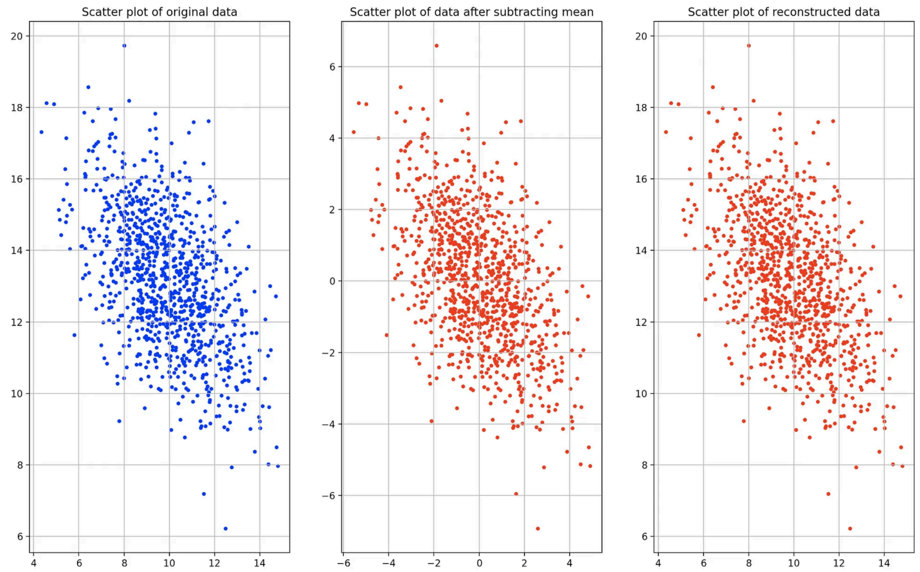
$Y = (X - \mu)\Gamma$으로 구한 PCs를 이용하여 다시 original data X로 reconstruct할 수 있음

$$reconX = Y\Gamma^T + \mu$$

$$= \Gamma^T(X - \mu)\Gamma + \mu$$

$$= (X - \mu)\Gamma^T\Gamma + \mu$$

$$= X - \mu + \mu = X$$



Shape of X : N x p
Shape of Γ : p x p, select top Eigenvector k : p x k
Shape of Y : N x k

Y와 Γ^T를 dot product해서 shape을 N x p로 reconstruct 가능

1. PCA

SVD와의 연관성

데이터 X에 대하여 covariance matrix C를 SVD로 분해하면

$C = U\Sigma U^T$로 분해할 수 있다. (Symmetric matrix의 svd이기 때문)

이때 U는 left singular vector of C 이고, Σ는 C의 singular values에 대한 diagonal matrix

그런데 symmetric matrix의 SVD에서 원래 matrix의 eigenvectors와 left singular vectors는 동일

Covariance matrix C는 symmetric하기 때문에 C의 SVD와 eigenvalue decomposition은 같은 결과를 도출함

columns of $U\Sigma$ -> principal component

=> PCA는 SVD의 special case

2.Asymptotic Properties of the PCs

2. Asymptotic Properties of PCs

Estimator of PC

실제로 PCA를 진행할 때는 sample data로 진행하는데, 이때 parameter들을 estimator로 바꿔서 계산해줘야 함

μ는 $\bar{x}$로, Σ는 S로, g_1은 S의 첫번째 eigenvector로 바꿔서 사용

first component $y_1 = (\mathcal{X} - 1_n\bar{x})g_1$

$\mathcal{S} = \mathcal{G}\mathcal{L}\mathcal{G}^T$로 분해할 수 있다고 할 때, principal component의 matrix form은 다음과 같음

$\mathcal{Y} = (\mathcal{X} - 1_n\bar{x}^T)\mathcal{G}$

2. Asymptotic Properties of PCs

Estimator of PC

$\mathscr{H} = \mathscr{I} - (n^{-1}1_n 1_n^T)$와 $\mathscr{H}1_n\bar{x}^T = 0$을 만족하는 centerting matrix $\mathscr{H}$이 있다고 가정하면
$\mathscr{Y}$에 대한 covariance matrix를 다음과 같이 나타낼 수 있다.

$$\begin{aligned} \mathcal{S}_{\mathcal{Y}} &= n^{-1}\mathcal{Y}^\top\mathcal{H}\mathcal{Y} = n^{-1}\mathcal{G}^\top(\mathcal{X} - 1_n\bar{x}^\top)^\top\mathcal{H}(\mathcal{X} - 1_n\bar{x}^\top)\mathcal{G} \\ &= n^{-1}\mathcal{G}^\top\mathcal{X}^\top\mathcal{H}\mathcal{X}\mathcal{G} = \mathcal{G}^\top\mathcal{S}\mathcal{G} = \mathcal{L} \end{aligned}$$

이때 $\mathscr{L} = diag(l_1, \ldots, l_p)$은 matrix of eigenvalues of $\mathscr{S}$ -> variance of y_i = igenvalue l_1

2. Asymptotic Properties of PCs

Sensitive to scale

PCA를 진행할 때 eigenvalue decomposition은 correlation matrix가 아닌 covariance matrix에서 시행

-> 데이터 scale에 민감해짐 (eigenvalue, eigenvector 전부 달라짐)

$$Cov(X,Y) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}),\ Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

$$\bar{x} = (214.9, 130.1, 129.9, 9.4, 10.6, 140.5)^\top .$$

$$\ell = (2.985, 0.931, 0.242, 0.194, 0.085, 0.035)^\top .$$

$$\mathcal{G} = \begin{pmatrix} -0.044 & 0.011 & 0.326 & 0.562 & -0.753 & 0.098 \\ 0.112 & 0.071 & 0.259 & 0.455 & 0.347 & -0.767 \\ 0.139 & 0.066 & 0.345 & 0.415 & 0.535 & 0.632 \\ 0.768 & -0.563 & 0.218 & -0.186 & -0.100 & -0.022 \\ 0.202 & 0.659 & 0.557 & -0.451 & -0.102 & -0.035 \\ -0.579 & -0.489 & 0.592 & -0.258 & 0.085 & -0.046 \end{pmatrix} .$$

2. Asymptotic Properties of PCs

Estimator of PC

X_1, X_2, X_3, X_6 에만 1/10을 곱해주고 eigenvalue, eigenvector를 구하면

$$\bar{x} = (21.49,\ 13.01,\ 12.99,\ 9.41,\ 10.65,\ 14.05)^\top.$$

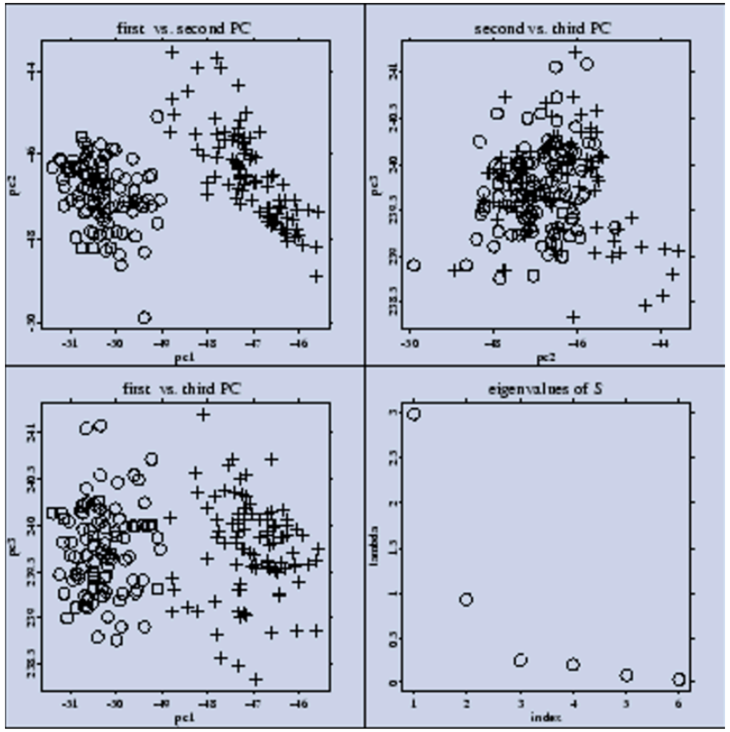
$$\ell = (2.101,\ 0.623,\ 0.005,\ 0.002,\ 0.001,\ 0.0004)^\top$$

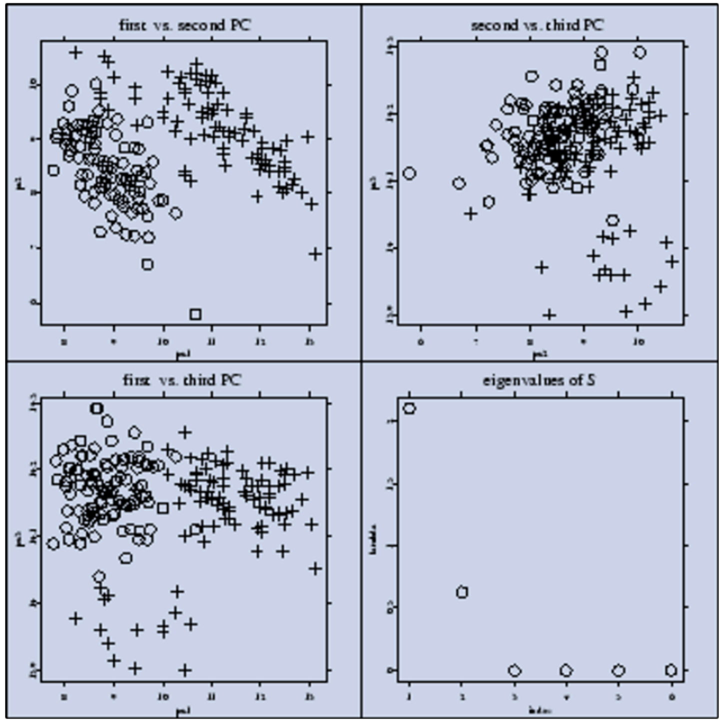
$$g_1 = (-0.005,\ 0.011,\ 0.014,\ 0.992,\ 0.113,\ -0.052)^\top$$

$$\mathcal{G} = \begin{pmatrix} -0.044 \\ 0.112 \\ 0.139 \\ 0.768 \\ 0.202 \\ -0.579 \end{pmatrix}$$

2. Asymptotic Properties of PCs

Estimator of PC





2. Asymptotic Properties of PCs

Asymptotic distribution of the sample PCs

THEOREM 9.4 Let $\Sigma > 0$ with distinct eigenvalues, and let $\mathcal{U} \sim m^{-1}W_p(\Sigma, m)$ with spectral decompositions $\Sigma = \Gamma\Lambda\Gamma^\top$, and $\mathcal{U} = \mathcal{G}\mathcal{L}\mathcal{G}^\top$. Then

(a)

$$\sqrt{m}(\ell - \lambda) \xrightarrow{\mathcal{L}} N_p(0, 2\Lambda^2),$$

where $\ell = (\ell_1, \ldots, \ell_p)^\top$ and $\lambda = (\lambda_1, \ldots, \lambda_p)^\top$ are the diagonals of $\mathcal{L}$ and Λ,

(b)

$$\sqrt{m}(g_j - \gamma_j) \xrightarrow{\mathcal{L}} N_p(0, \mathcal{V}_j),$$

with $\mathcal{V}_j = \lambda_j \sum_{k \neq j} \frac{\lambda_k}{(\lambda_k - \lambda_j)^2} \gamma_k \gamma_k^\top$,

(c)

$$Cov(g_j, g_k) = \mathcal{V}_{jk},$$

where the (r,s)-element of the matrix $\mathcal{V}_{jk}(p \times p)$ is $-\frac{\lambda_j \lambda_k \gamma_{rk} \gamma_{sj}}{[m(\lambda_j - \lambda_k)^2]}$,

(d)

the elements in ℓ are asymptotically independent of the elements in $\mathcal{G}$.

ESC

2. Asymptotic Properties of PCs

Asymptotic distribution of the sample PCs

Example)

$n\mathcal{S} \sim m^{-1}W_p(\Sigma, n-1)$이고, $X_1, \ldots, X_n$는 $N(\mu, \Sigma)$에서 뽑힌 샘플이라고 했을 때,

$$\sqrt{n-1}(\ell_j - \lambda_j) \xrightarrow{\mathcal{L}} N(0, 2\lambda_j^2), \quad j = 1, \ldots, p.$$ 가 성립 By a

Right hand side의 variance가 λ_j에 관한 식이기 때문에 이를 제거하기 위해 log transformation 사용

$f(l_j) = log(l_j)$라고 두고, $\frac{d}{dl_j} f\,|_{l_j=\lambda_j} = \frac{1}{\lambda_j}$ 를 이용

2. Asymptotic Properties of PCs

Asymptotic distribution of the sample PCs

THEOREM 4.11 If $\sqrt{n}(t-\mu) \xrightarrow{\mathcal{L}} N_p(0, \Sigma)$ and if $f = (f_1, \ldots, f_q)^\top : \mathbb{R}^p \to \mathbb{R}^q$ are real valued functions which are differentiable at $\mu \in \mathbb{R}^p$, then $f(t)$ is asymptotically normal with mean $f(\mu)$ and covariance $\mathcal{D}^\top \Sigma \mathcal{D}$, i.e.,

$$\sqrt{n}\{f(t) - f(\mu)\} \xrightarrow{\mathcal{L}} N_q(0, \mathcal{D}^\top \Sigma \mathcal{D}) \qquad \text{for} \quad n \longrightarrow \infty, \qquad (4.56)$$

Theorem 4.11에 의해 아래처럼 변환 가능하다.

$$\sqrt{n-1}(\log \ell_j - \log \lambda_j) \longrightarrow N(0, 2).$$

$$\sqrt{\frac{n-1}{2}}(\log \ell_j - \log \lambda_j) \xrightarrow{\mathcal{L}} N(0, 1)$$

2. Asymptotic Properties of PCs

Asymptotic distribution of the sample PCs

그렇게 구한 식에서 $log\lambda_j$에 대한 two-sided confidence interval을 $1-\alpha = 0.95$ 신뢰수준으로 계산하면

$$\log(\ell_j) - 1.96\sqrt{\frac{2}{n-1}} \le \log\lambda_j \le \log(\ell_j) + 1.96\sqrt{\frac{2}{n-1}}.$$

$$\ell_1 = 2.98. \qquad \log(2.98) \pm 1.96\sqrt{\frac{2}{199}} = \log(2.98) \pm 0.1965. \qquad P\{\lambda_1 \in (2.448, 3.62)\} \approx 0.95.$$

2. Asymptotic Properties of PCs

Variance explained by first q PCs

$$\psi = \frac{\lambda_1 + \cdots + \lambda_q}{\sum_{j=1}^{p} \lambda_j}.$$

$$\widehat{\psi} = \frac{\ell_1 + \cdots + \ell_q}{\sum_{j=1}^{p} \ell_j}.$$

In practice , λ estimated by l

아까 Theorem 9.4를 통해 $\sqrt{n-1}(l-\lambda)$ distribution을 알고 있지만,

ψ는 nonlinear한 function이기 때문에 transformation해줘야 함

$$\sqrt{n-1}(\widehat{\psi} - \psi) \xrightarrow{\mathcal{L}} N(0, \mathcal{D}^\top \mathcal{V} \mathcal{D})$$

2. Asymptotic Properties of PCs

Variance explained by first q PCs

$$\sqrt{n-1}(\widehat{\psi}-\psi) \xrightarrow{\mathcal{L}} N(0, \mathcal{D}^\top \mathcal{V} \mathcal{D}) \qquad \mathcal{V} = 2\Lambda^2 \qquad \mathcal{D} = (d_1, \dots, d_p)^\top$$

$$d_j = \frac{\partial \psi}{\partial \lambda_j} = \begin{cases} \dfrac{1-\psi}{\operatorname{tr}(\Sigma)} & \text{for } 1 \le j \le q, \\ \dfrac{-\psi}{\operatorname{tr}(\Sigma)} & \text{for } q+1 \le j \le p. \end{cases}$$

2. Asymptotic Properties of PCs

Variance explained by first q PCs

이렇게 나타낸 분포 수렴식을 정리하면

$$\sqrt{n-1}(\widehat{\psi}-\psi) \xrightarrow{\mathcal{L}} N(0,\omega^2),$$

$$\begin{aligned}\omega^2 &= \mathcal{D}^\top \mathcal{V} \mathcal{D} = \frac{2}{\{\operatorname{tr}(\Sigma)\}^2}\{(1-\psi)^2(\lambda_1^2+\cdots+\lambda_q^2)+\psi^2(\lambda_{q+1}^2+\cdots+\lambda_p^2)\} \\ &= \frac{2\operatorname{tr}(\Sigma^2)}{\{\operatorname{tr}(\Sigma)\}^2}(\psi^2-2\beta\psi+\beta)\end{aligned}$$

$$\beta = \frac{\lambda_1^2+\cdots+\lambda_q^2}{\lambda_1^2+\cdots+\lambda_p^2}.$$

2. Asymptotic Properties of PCs

Variance explained by first q PCs

Example)

Swiss bank data을 통해 PCA를 한 결과 first PC만으로 전체 분산의 67%를 차지한다고 알려져있는데, 75%를 차지하는지에 대한 테스트

$$
\begin{aligned}
\widehat{\beta} &= \frac{\ell_1^2}{\ell_1^2 + \cdots + \ell_p^2} = \frac{(2.985)^2}{(2.985)^2 + (0.931)^2 + \cdots (0.035)^2} = 0.902 \\
\mathrm{tr}(\mathcal{S}) &= 4.472 \\
\mathrm{tr}(\mathcal{S}^2) &= \sum_{j=1}^{p} \ell_j^2 = 9.883 \\
\widehat{\omega}^2 &= \frac{2\,\mathrm{tr}(\mathcal{S}^2)}{\{\mathrm{tr}(\mathcal{S})\}^2}(\widehat{\psi}^2 - 2\widehat{\beta}\widehat{\psi} + \widehat{\beta}) \\
&= \frac{2 \cdot 9.883}{(4.472)^2}\{(0.668)^2 - 2(0.902)(0.668) + 0.902\} = 0.142.
\end{aligned}
$$

$$0.668 \pm 1.96\sqrt{\frac{0.142}{199}} = (0.615, 0.720).$$

Interpretation of the Principal Components

Interpretation of the Principal Components

Data reduction, Eigenvalue

PCA의 주요 목적. Data reduction, Dimension reduction

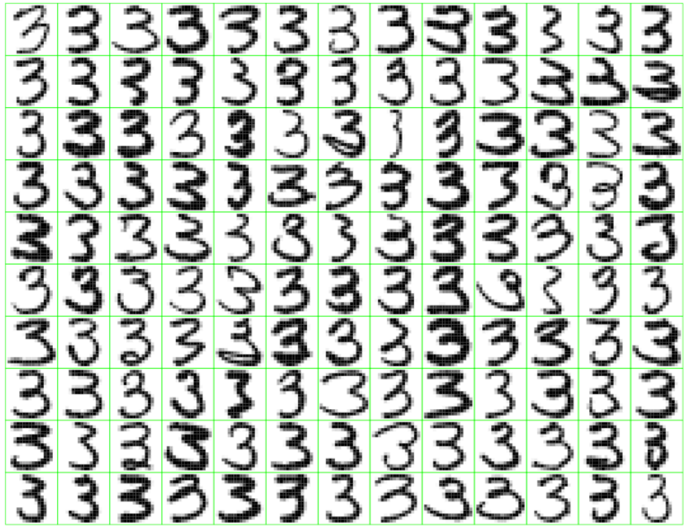
p-dimension → k-dimension

데이터를 가장 잘 설명하는 k개의 축을 뽑아서, 적은 수의 feature만으로 효율적으로 설명하는 것이 주요 목적.

Interpretation of the Principal Components

Example: Handwritten Digit
(*Elements of Statistical Learning* by Hastie, Tibshirani, Friedman)

- A sample of 130 handwritten 3s, each a digitized 16×16 grayscale image, from a total of 658 such 3s, shows considerable variation in writing styles, character thickness and orientation.

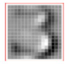
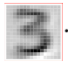


- These images can be considered as points x_i in $\mathbb{R}^{256}$, and their principal components can be computed.

- These images can be considered as points x_i in $\mathbb{R}^{256}$, and their principal components can be computed.

- The two-component model:

$$\hat{f}(\lambda) = \bar{x} + \lambda_1 v_1 + \lambda_2 v_2$$

$=$  $+ \lambda_1 \cdot$  $+ \lambda_2 \cdot$ .

where λ_1, λ_2 are PC scores.

- PC_1 mainly accounts for the lengthening of the lower tail of the three, while PC_2 accounts for character thickness.

Interpretation of the Principal Components

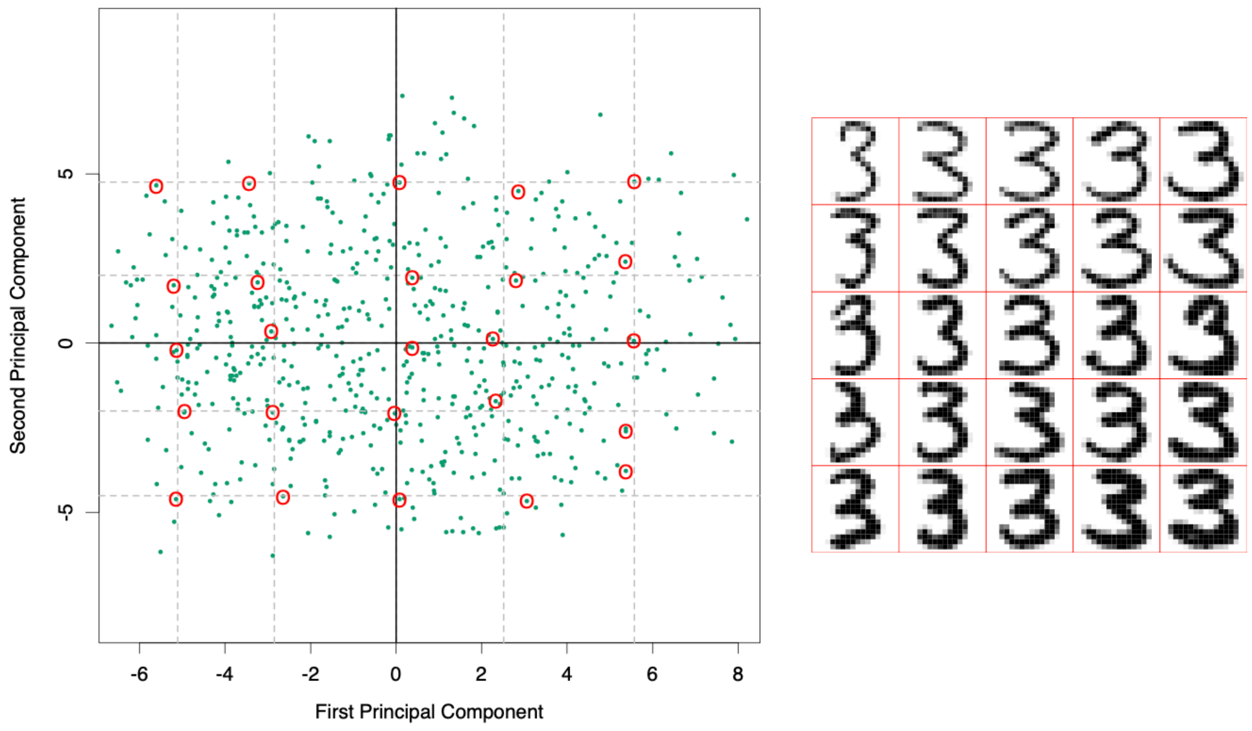


FIGURE 14.23. *(Left panel:) the first two principal components of the hand-written threes. The circled points are the closest projected images to the vertices of a grid, defined by the marginal quantiles of the principal components. (Right panel:) The images corresponding to the circled points. These show the nature of the first two principal components.*

Interpretation of the Principal Components

Data reduction, Eigenvalue

k 값을 어떻게 정할 것인가? 몇 개의 component를 가져갈 것인가

The Number of Principal Components

- The amount of total sample variance explained
  - λ_i : ith component Y_i로 설명되는 Variability
  - e.g. retaining enough PCs to explain, say, 90% of the total variation
- The relative sizes of the eigenvalues
  - e.g. retaining PCs where the eigenvalue is above the average
  - scree plot

Interpretation of the Principal Components

Data reduction, Eigenvalue

elbow인 $i = 2$ 또는 $i = 3$으로 선택. Without other any evidence.

해당 지점에서 sample principal components가 효과적으로

total sample variance를 요약한다고 본다.

그 외에도, 의미 있다고 생각되는 Principal Component를 채택하고는 한다.

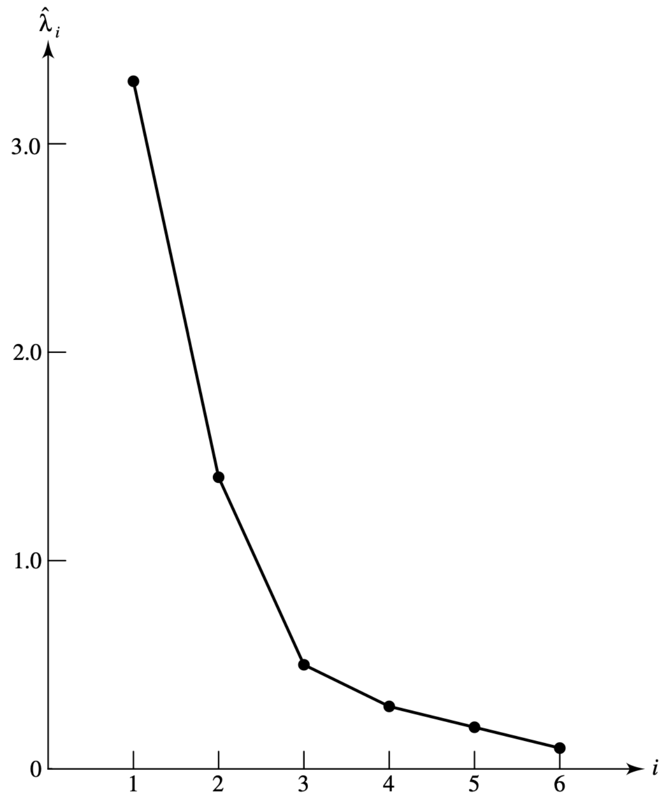


Figure 8.2 A scree plot.

Interpretation of the Principal Components

+ Smallest Eigenvalue

큰 $\lambda_1, \lambda_2, \ldots$ 만 중요하게 여길게 아니라, smallest eigenvalue도 눈여겨 볼 필요가 존재한다.

Smallest $\lambda_p \rightarrow 0$

linear dependency, multicollinearity 존재 의미

determinant가 0에 수렴, by rounding error

종종 Condition Number가 특정 숫자보다 크면 multicollinearity의 존재로 봄

이렇게 multicollinearity가 존재할 때, PCA를 이용하면 아주 효과적이다.

highly correlated인 feature들을 principal component로 묶고, 각 principal component 끼리는 ***uncorrelated***

다른 분석들을 행하기 전에 principal component로 바꿔주면,

이 Principal Component 사이에서는 multicollinearity가 거의 해소됐기에 아주 효과적

Interpretation of the Principal Components

Relationship between Y_i and X_k

ith Principal Component Y_i에 대한 kth feature X_k의 기여도 측정

1. Coefficient

e_{ik} ; relation between Y_i and X_k, given $\mathbf{X}$

$\hat{e}_{ik}$; sample ver.

2. Correlation

ρ_{Y_i,X_k} ; only between Y_i and X_k, not $\mathbf{X}$

e_{ik} is proportional to the correlation coefficient between Y_i and X_k

$$\rho_{Y_i,X_k} = \frac{e_{ik}\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}}$$

$r_{\hat{y}_i,x_k}$; sample ver.

$$r_{\hat{y}_i,x_k} = \frac{\hat{e}_{ik}\sqrt{\hat{\lambda}_i}}{\sqrt{s_{kk}}}$$

Interpretation of the Principal Components

Relationship between Y_i and X_k

Coef.

$\mathbf{X}$를 모두 반영함

$Var(X_k)$가 크면 Coef.도 크게 나오는 경향 존재

Corr.

$Var(X_k)$와 관계 없음. Standardized

오직 X_k와의 관계이기에, $\mathbf{X}$를 구성하는 다른 feature들과의 상관관계를 고려하지 못함

$\Rightarrow$ both should be examined, and standardize will affect

Interpretation of the Principal Components

Relationship between Y_i and X_k

Example 8.2 in Johnson and Wichern - Applied Multivariate Statistical Analysis

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

The eigenvalue–eigenvector pairs from $\mathbf{\Sigma}$ are

$$\begin{aligned} \lambda_1 &= 100.16, & \mathbf{e}_1' &= [.040, .999] \\ \lambda_2 &= \quad .84, & \mathbf{e}_2' &= [.999, -.040] \end{aligned}$$

Similarly, the eigenvalue–eigenvector pairs from $\boldsymbol{\rho}$ are

$$\begin{aligned} \lambda_1 &= 1 + \rho = 1.4, & \mathbf{e}_1' &= [.707, .707] \\ \lambda_2 &= 1 - \rho = \ .6, & \mathbf{e}_2' &= [.707, -.70 \end{aligned}$$

The respective principal components become

$$\mathbf{\Sigma}: \quad \begin{aligned} Y_1 &= .040X_1 + .999X_2 \\ Y_2 &= .999X_1 - .040X_2 \end{aligned}$$

and

$$\boldsymbol{\rho}: \quad \begin{aligned} Y_1 &= .707Z_1 + .707Z_2 = .707\left(\frac{X_1 - \mu_1}{1}\right) + .707\left(\frac{X_2 - \mu_2}{10}\right) \\ &= .707\,(X_1 - \mu_1) + .0707(X_2 - \mu_2) \\ Y_2 &= .707Z_1 - .707Z_2 = .707\left(\frac{X_1 - \mu_1}{1}\right) - .707\left(\frac{X_2 - \mu_2}{10}\right) \\ &= .707\,(X_1 - \mu_1) - .0707(X_2 - \mu_2) \end{aligned}$$

Monitoring Quality with Principal Components

Monitoring Quality with Principal Components

Checking a Given Set of Measurements for Stabiliity

Can reveal suspect observations.

With first few Principal Components

Using ellipse format chart

First two sample principal components만 하는 이유

feature가 10개만 되어도, 45가지 경우의 수

나머지는 다른 방법(T^2-Chart) 이용할 예정

특히 공장처럼, 지속적으로 안정적인 데이터가 만들어지는 곳.

Observ.의 stability를 알아보는 방법 : Ellipse Chart

Monitoring Quality with Principal Components

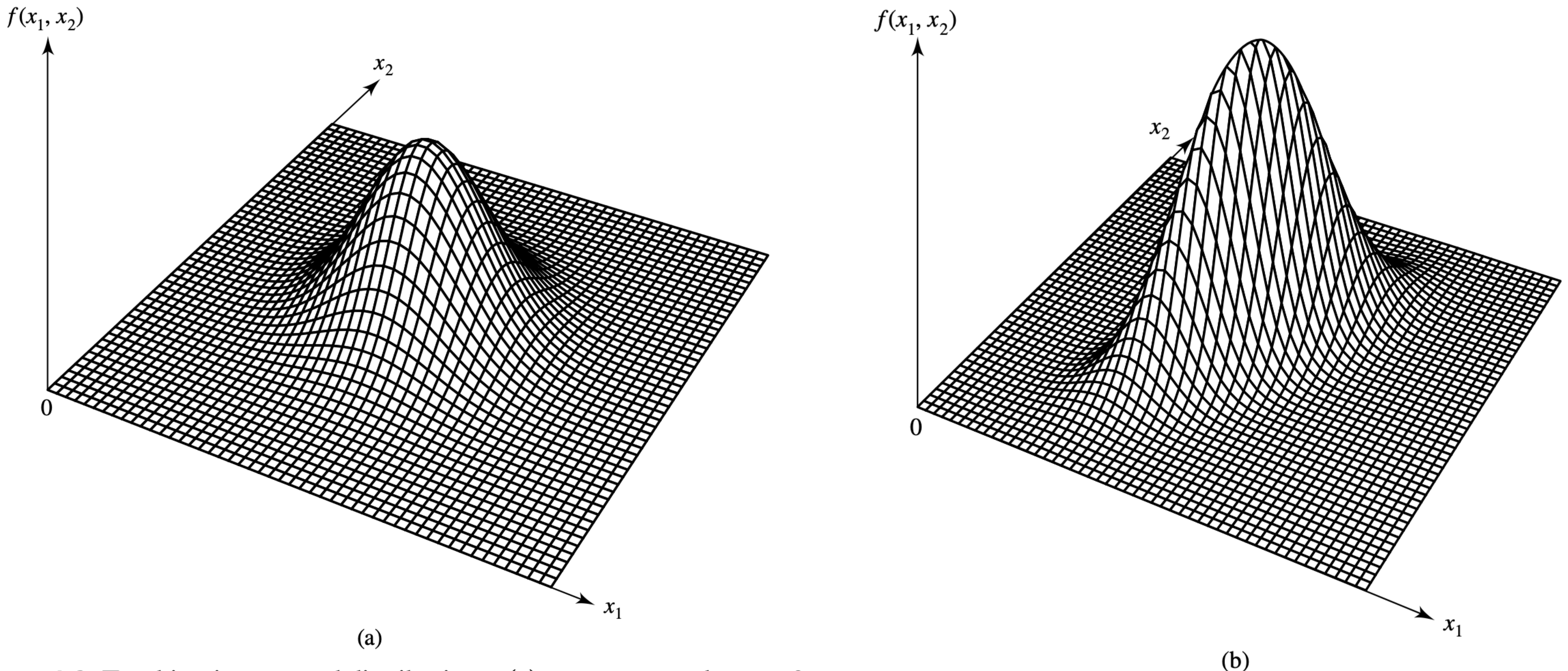


Figure 4.2 Two bivariate normal distributions. (a) $\sigma_{11} = \sigma_{22}$ and $\rho_{12} = 0$. (b) $\sigma_{11} = \sigma_{22}$ and $\rho_{12} = .75$.

Monitoring Quality with Principal Components

Ellipse Chart

Contours of constant density for the p-dimensional normal distribution are ellipsoids defined by $\mathbf{x}$ such the that

$$(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = c^2 \tag{4-7}$$

These ellipsoids are centered at $\boldsymbol{\mu}$ and have axes $\pm c\sqrt{\lambda_i}\,\mathbf{e}_i$, where $\boldsymbol{\Sigma}\mathbf{e}_i = \lambda_i\mathbf{e}_i$ for $i = 1, 2, \ldots, p$.

The solid ellipsoid of $\mathbf{x}$ values satisfying

$$(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \leq \chi_p^2(\alpha) \tag{4-8}$$

has probability $1 - \alpha$.

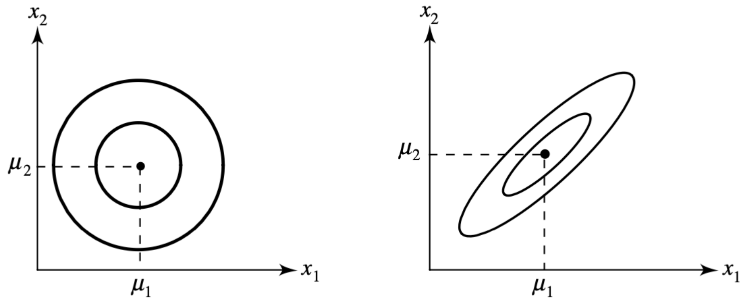


Figure 4.4 The 50% and 90% contours for the bivariate normal distributions in Figure 4.2.

Monitoring Quality with Principal Components

Ellipse Chart

It is informative to consider principal components derived from multivariate normal random variables. Suppose $\mathbf{X}$ is distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. We know from (4-7) that the density of $\mathbf{X}$ is constant on the $\boldsymbol{\mu}$ centered ellipsoids

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

For Convenience, set $\mu = \mathbf{0}$

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

$y_1 = \mathbf{e}_1'\mathbf{x}$, $y_2 = \mathbf{e}_2'\mathbf{x}, \ldots, y_p = \mathbf{e}_p'\mathbf{x}$, we have

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

Monitoring Quality with Principal Components

Ellipse Chart

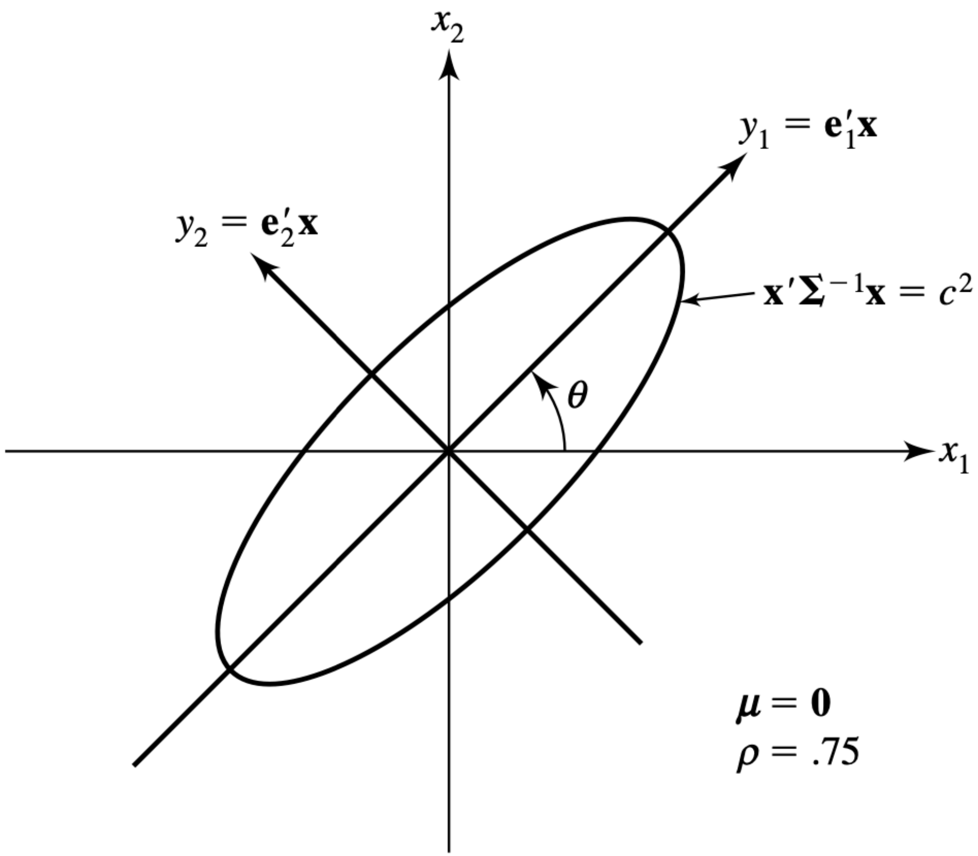


Figure 8.1 The constant density ellipse $\mathbf{x}'\mathbf{\Sigma}^{-1}\mathbf{x} = c^2$ and the principal components y_1, y_2 for a bivariate normal random vector $\mathbf{X}$ having mean $\mathbf{0}$.

Monitoring Quality with Principal Components

Checking a Given Set of Measurements for Stabiliity

For sample data,

suppose the underlying distribution of X is nearly $N_p(\mu, \Sigma)$

from the sample values $\mathbf{x_j}$, we can approximate μ by $\bar{\mathbf{x}}$ and Σ by $\mathbf{S}$

even when the normal assumption is suspect and the scatter plot may depart somewhat from an elliptical pattern

$$(\mathbf{x} - \bar{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x} - \bar{\mathbf{x}}) = c^2$$

Monitoring Quality with Principal Components

Checking a Given Set of Measurements for Stabiliity

For visualizing, select two axes $\hat{y}_1, \hat{y}_2$

- $\hat{y}_3, \ldots, \hat{y}_p$ 버리고 그래프 그리기
- 꼭 $\hat{y}_1, \hat{y}_2$일 필요는 없다. 필요한 축 선택

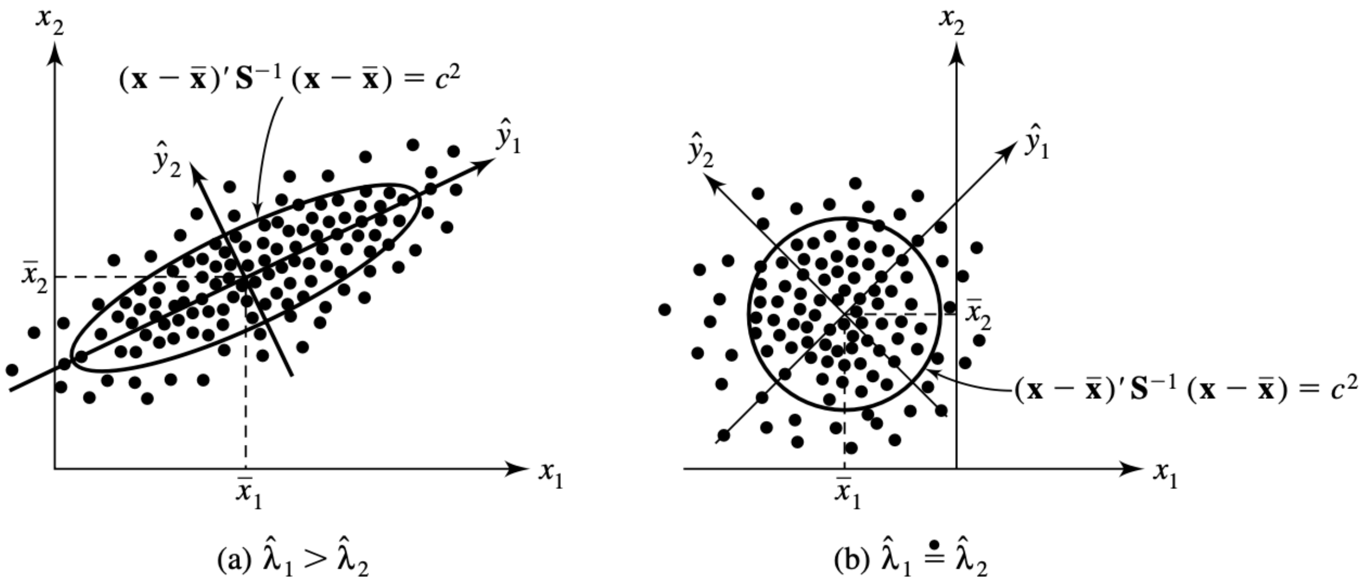


Figure 8.4 Sample principal components and ellipses of constant distance.

(a) $\hat{y}_1$채택하면 좋을 것으로 보임
(b) circular, 무엇 하나를 버리기 곤란하다.

Monitoring Quality with Principal Components

Checking a Given Set of Measurements for Stabiliity

Construct an ellipse format chart for the pairs of values $(\hat{y}_{j1}, \hat{y}_{j2})$ for $j = 1,2,\ldots,n$

Confidence Ellipse : $\dfrac{\hat{y}_1^{\,2}}{\hat{\lambda}_1} + \dfrac{\hat{y}_2^{\,2}}{\hat{\lambda}_2} \le \chi_2^2(\alpha)$

Each $\mathbf{X}_j - \overline{\mathbf{X}}$ has a normal distribution but, $\mathbf{X}_j - \overline{\mathbf{X}}$ is not independent of the sample covariance matrix $\mathbf{S}$. However to set control limits, we approximate that $(\mathbf{X}_j - \overline{\mathbf{X}})'\mathbf{S}^{-1}(\mathbf{X}_j - \overline{\mathbf{X}})$ has a chi-square distribution.

Ellipse Format Chart. The ellipse format chart for a bivariate control region is the more intuitive of the charts, but its approach is limited to two variables. The two characteristics on the jth unit are plotted as a pair (x_{j1}, x_{j2}). The 95% quality ellipse consists of all $\mathbf{x}$ that satisfy

$$(\mathbf{x} - \bar{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x} - \bar{\mathbf{x}}) \le \chi_2^2(.05) \qquad (5\text{-}32)$$

Monitoring Quality with Principal Components

Table 5.8 Five Types of Overtime Hours for the Madison, Wisconsin, Police Department

x_1 Legal Appearances Hours	x_2 Extraordinary Event Hours	x_3 Holdover Hours	x_4 COA[1] Hours	x_5 Meeting Hours
3387	2200	1181	14,861	236
3109	875	3532	11,367	310
2670	957	2502	13,329	1182
3125	1758	4510	12,328	1208
3469	868	3032	12,847	1385
3120	398	2130	13,979	1053
3671	1603	1982	13,528	1046
4531	523	4675	12,699	1100
3678	2034	2354	13,534	1349
3238	1136	4606	11,609	1150
3135	5326	3044	14,189	1216
5217	1658	3340	15,052	660
3728	1945	2111	12,236	299
3506	344	1291	15,482	206
3824	807	1365	14,900	239
3516	1223	1175	15,078	161

[1] Compensatory overtime allowed.

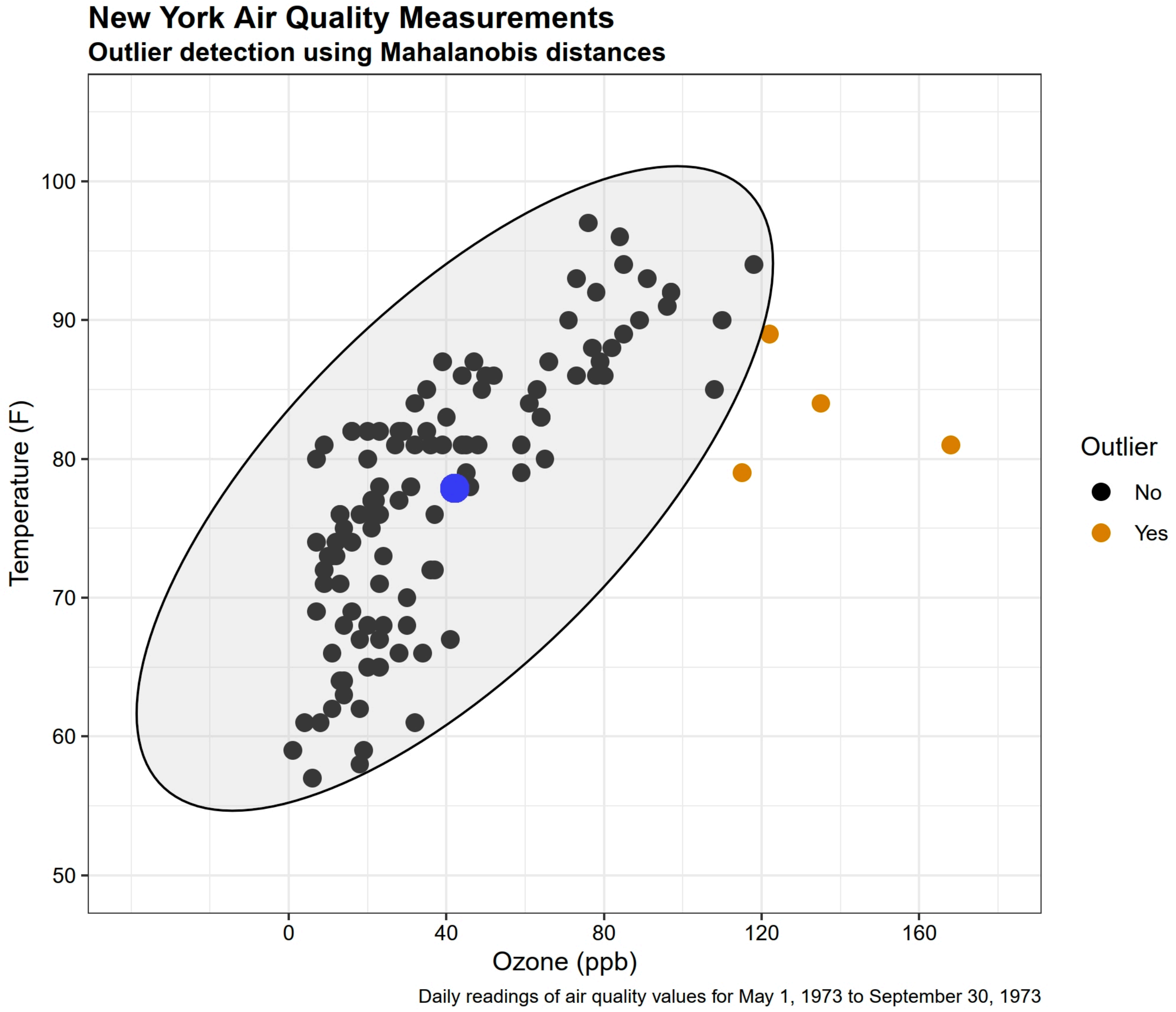
Table 8.1 Eigenvectors and Eigenvalues of Sample Covariance Matrix for Police Department Data

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

Table 8.2 Values of the Principal Components for the Police Department Data

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

Monitoring Quality with Principal Components



Monitoring Quality with Principal Components

Checking a Given Set of Measurements for Stabiliity

With remaining latter Principal Components

with Principal Componentss not involved in the ellipse format chart

T^2 chart

Based on the last $p-2$ components,

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

→ Sum of $p-2$ chi-square r.v.s

Monitoring Quality with Principal Components

Checking a Given Set of Measurements for Stabiliity

Sample data로 가면

$$T_j^2 = \frac{\hat{y}_{j3}^2}{\hat{\lambda}_3} + \frac{\hat{y}_{j4}^2}{\hat{\lambda}_4} + \cdots + \frac{\hat{y}_{jp}^2}{\hat{\lambda}_p}$$

$$UCL = \chi^2_{p-2}(\alpha)$$

Large sample approximation하면 chi-square dist.

Monitoring Quality with Principal Components

Checking a Given Set of Measurements for Stabiliity

$$T_j^2 = \frac{\hat{y}_{j3}^2}{\hat{\lambda}_3} + \frac{\hat{y}_{j4}^2}{\hat{\lambda}_4} + \frac{\hat{y}_{j5}^2}{\hat{\lambda}_5}$$

where the first value is $T^2 = .891$ and so on. The T^2-chart is shown in Figure 8.8.

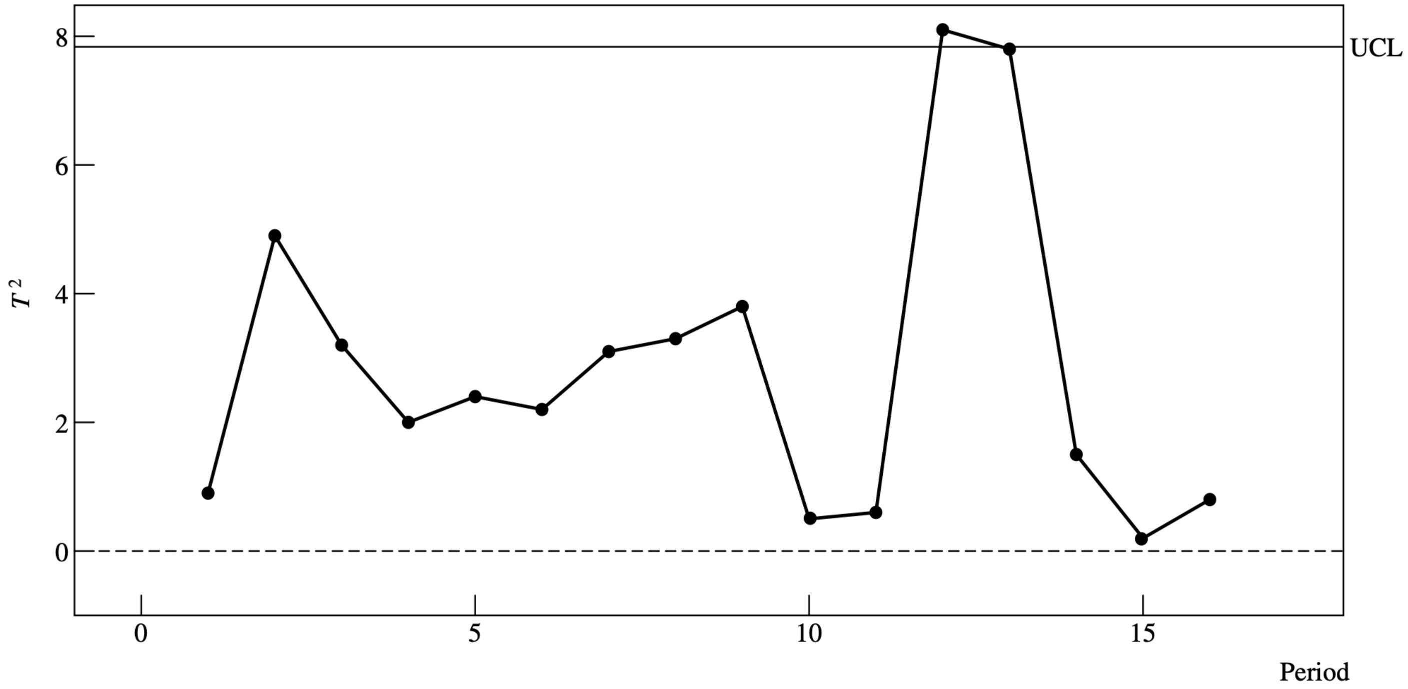


Figure 8.8 A T^2-chart based on the last three principal components of overtime hours.

Period 11 이후 뭔가 발생. First two PC로 설명 안되는 무언가.
Extraordinary event 피크 찍은 뒤에 시간 조정에 들어갔나? 의심.

Monitoring Quality with Principal Components

Checking a Given Set of Measurements for Stabiliity

The first components allow the detection of outliers which inflate the variances and covariances. These outliers are also extreme on original variables, so they can be directly detected.

On the other hand, outliers not visible on original variables (those that perturb the correlation between variables) will be detected on the last principal components.

Monitoring Quality with Principal Components

Controlling Future Values

ellipse chart를 보고, Observ. 11을 빼기로 결정.

Principal Components에 변화 발생. 특히 순서

Table 8.3 Eigenvectors and Eigenvalues from the 15 Stable Observations

	$\hat{\mathbf{e}}_1$	$\hat{\mathbf{e}}_2$	$\hat{\mathbf{e}}_3$	$\hat{\mathbf{e}}_4$	$\hat{\mathbf{e}}_5$
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

Extraordinary event overtime이 엄청 컸던
observ.를 빼니까, 그걸 설명하던 기존의 PC2가 영향력이 낮아진 것으로 예상.

" 앞으로는 stable operation of the process에서 observ.를 얻고 싶다 "
→ limiting future values

Monitoring Quality with Principal Components

Controlling Future Values

With remaining latter Principal Components,

전과 비슷하게 T^2- chart 적용

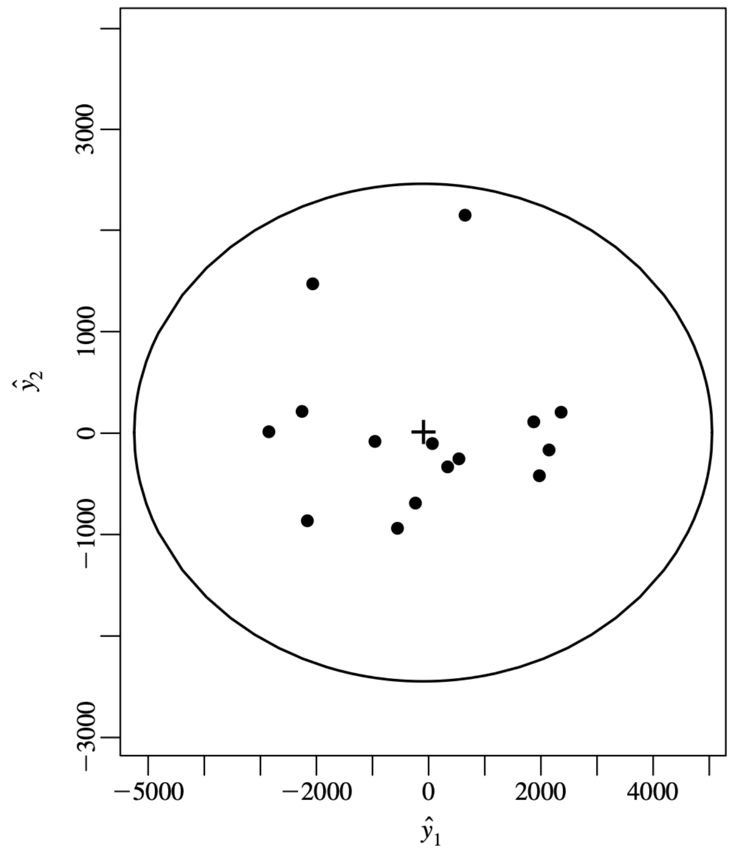


Figure 8.9 A 99% ellipse format chart for the first two principal components of future values of overtime.

Example
2019-20 Premier League

Example ; 2019-20 Premier League

Team	W	D	L	G	GA	GD
Liverpool	32	3	3	85	33	52
Manchester City	26	3	9	102	35	67
Manchester United	18	12	8	66	36	30
Chelsea	20	6	12	69	54	15
Leicester City	18	8	12	67	41	26
Tottenham Hotspur	16	11	11	61	47	14
Wolverhampton	15	14	9	51	40	11
Arsenal	14	14	10	56	48	8
Sheffield United	14	12	12	39	39	0
Burnley	15	9	14	43	50	-7

Example ; 2019-20 Premier League

$$\bar{\mathbf{x}} = \begin{pmatrix} 14.4 \\ 9.2 \\ 14.4 \\ 51.7 \\ 51.7 \\ 0 \end{pmatrix} \qquad \mathbf{S} = \begin{pmatrix} 38.3 & -9.18 & -29.2 & 103 & -57 & 160 \\ -9.18 & 10.2 & -0.98 & -27.5 & -2.24 & -25.2 \\ -29.2 & -0.98 & 30.1 & -75.3 & 59.3 & -135 \\ 103 & -27.5 & -75.3 & 336 & -147 & 483 \\ -57 & -2.24 & 59.3 & -147 & 134 & -281 \\ 160 & -25.2 & -135 & 483 & -281 & 764 \end{pmatrix}$$

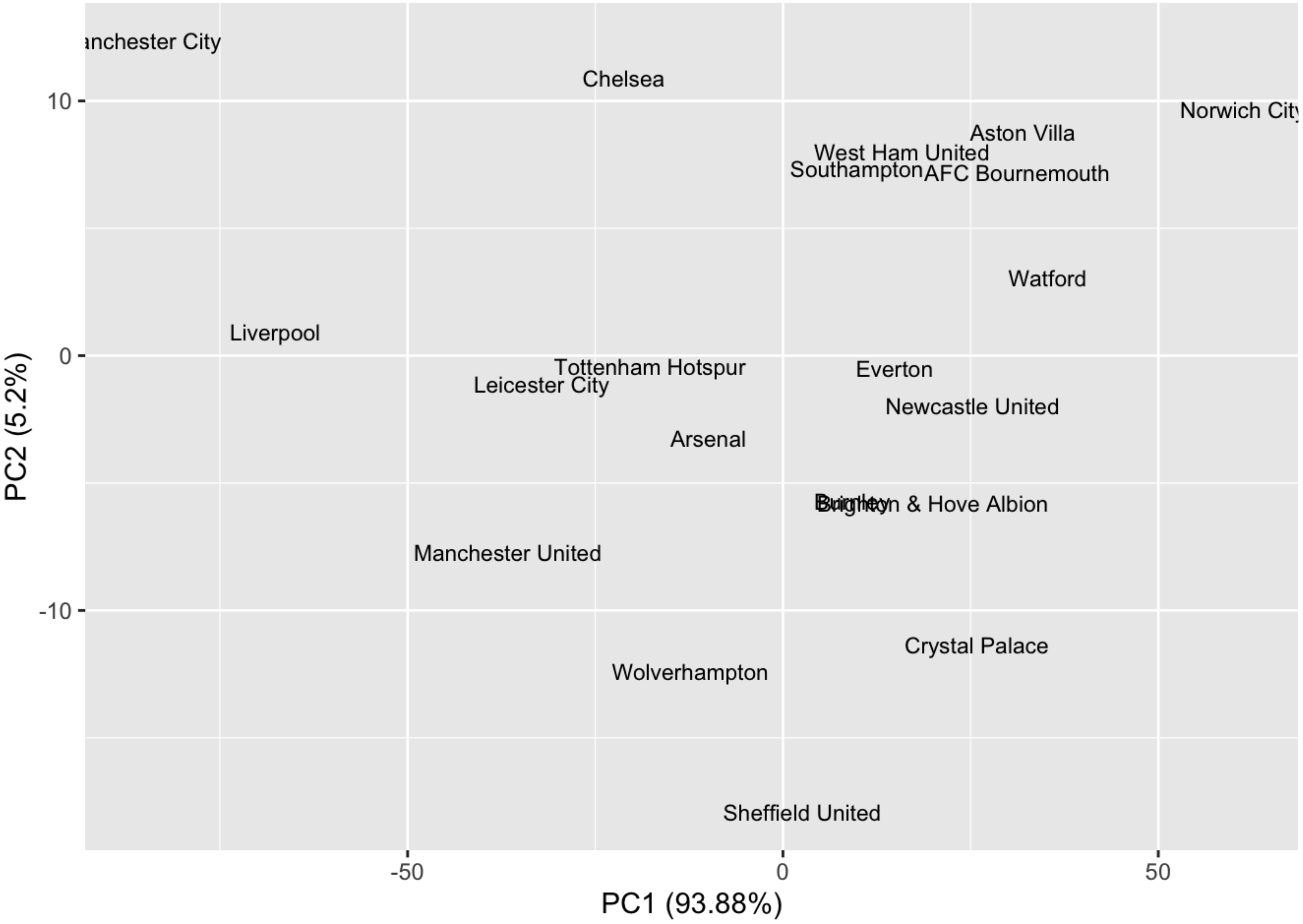
$$\mathbf{\Lambda} = \operatorname{diag}\begin{pmatrix} 1300 & 71.9 & 8.05 & 4.62 & -2.65e-14 & -3.73e-14 \end{pmatrix}$$

$$\begin{pmatrix} 0.939 & 0.052 & 0.00583 & 0.00334 & -1.92e-17 & -2.7e-17 \end{pmatrix}$$

$$\mathbf{V} = [\mathbf{v}_1 \ldots \mathbf{v}_6] = \begin{pmatrix} -0.166 & 0.0262 & -0.707 & 0.373 & 0.222 & -0.533 \\ 0.0282 & -0.275 & 0.661 & 0.391 & 0.222 & -0.533 \\ 0.138 & 0.249 & 0.0455 & -0.764 & 0.222 & -0.533 \\ -0.502 & 0.6 & 0.202 & 0.117 & 0.533 & 0.222 \\ 0.285 & 0.701 & 0.11 & 0.286 & -0.533 & -0.222 \\ -0.787 & -0.101 & 0.0915 & -0.169 & -0.533 & -0.222 \end{pmatrix}$$

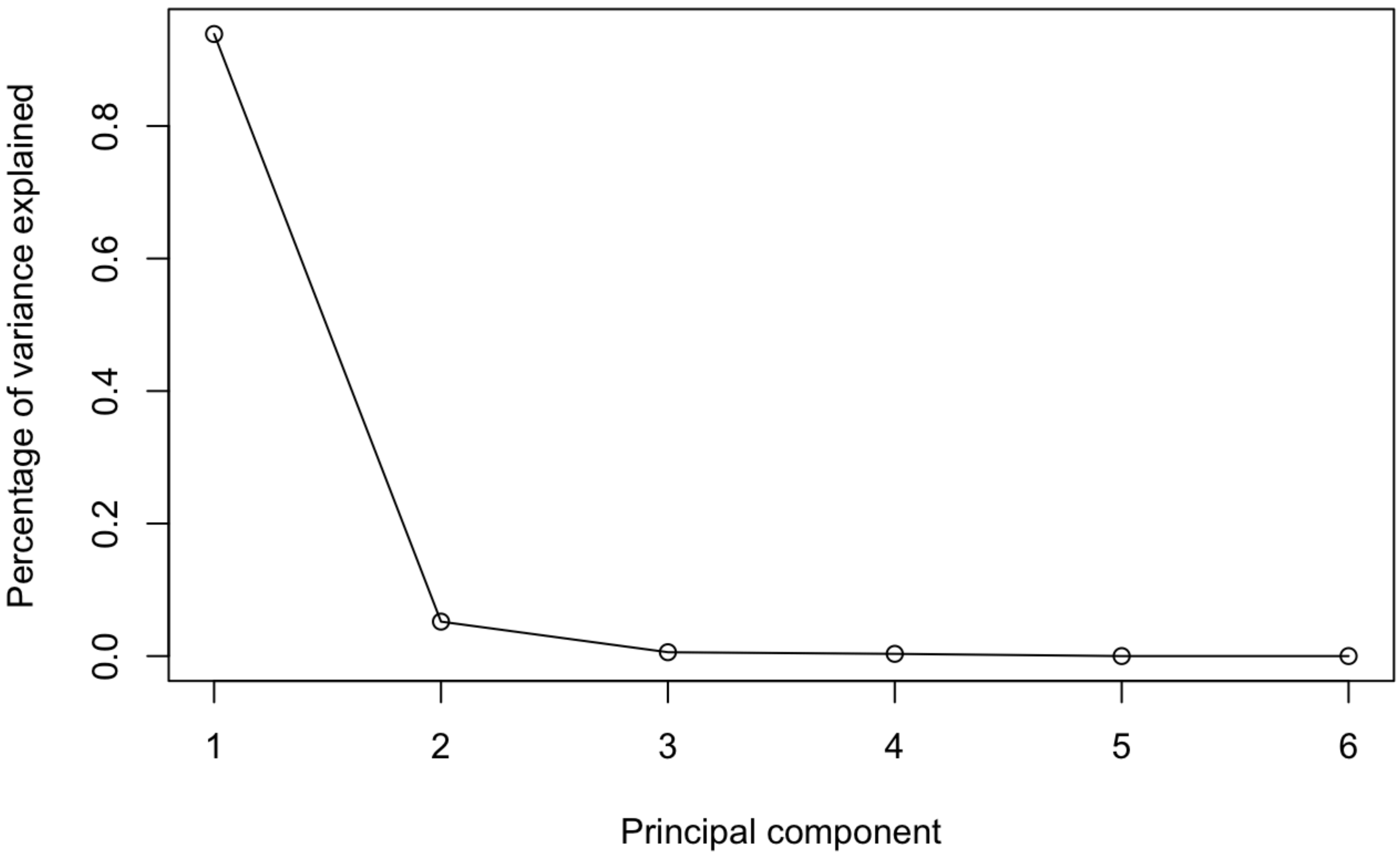
$$\begin{aligned} y_{i1} = {} & -0.17(W_i - \bar{W}) + 0.03(D_i - \bar{D}) + 0.14(L_i - \bar{L}) \\ & + -0.5(G_i - \bar{G}) + 0.28(GA_i - \bar{GA}) + -0.79(GD_i - \bar{GD}) \end{aligned}$$

Example ; 2019-20 Premier League



Team	PC1	PC2
Liverpool	-67.6	0.9
Manchester City	-85.6	12.3
Manchester United	-36.7	-7.7
Chelsea	-21.2	10.9
Leicester City	-32.2	-1.1

Example ; 2019-20 Premier League



Use **R** instead of **S**

scale이 너무 다르면 **R**로 바꿔주는게 더 좋다

그러나 큰 차이가 없고,

서로 다른 scale만큼 중요도의 차이가 난다면

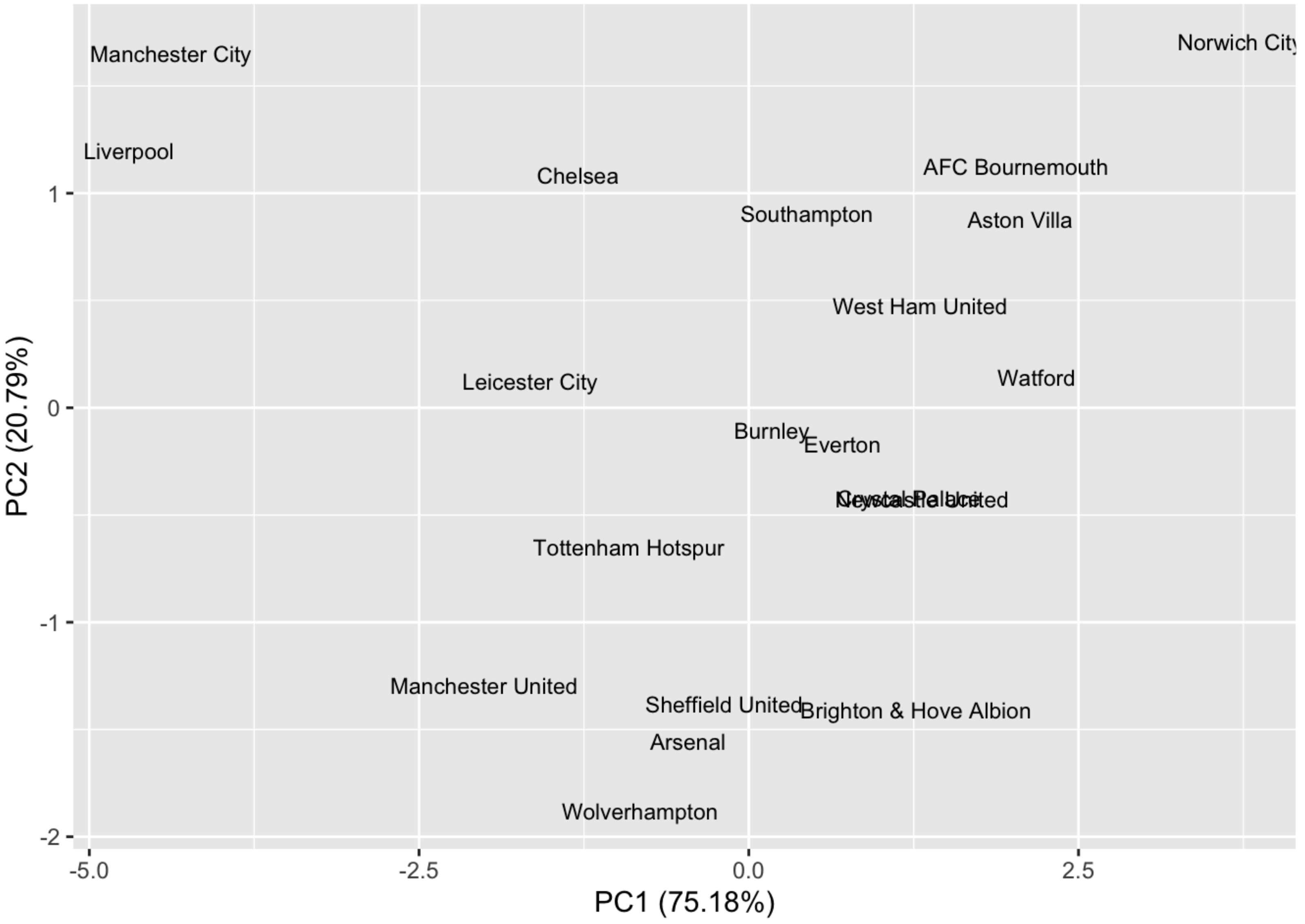
안바꿔주는게 더 좋을 것.

Example ; 2019-20 Premier League

$$\mathbf{\Lambda} = \mathrm{diag}\,(\,4.51 \quad 1.25 \quad 0.156 \quad 0.0863 \quad 3.68e-32 \quad 2.48e-33\,)$$

$$\mathbf{V} = [\mathbf{v}_1 \ldots \mathbf{v}_6] = \begin{pmatrix} -0.456 & 0.149 & -0.342 & -0.406 & 0.466 & 0.52 \\ 0.143 & -0.844 & 0.344 & -0.143 & 0.24 & 0.268 \\ 0.432 & 0.321 & 0.186 & 0.541 & 0.413 & 0.461 \\ -0.438 & 0.214 & 0.7 & -0.0181 & 0.389 & -0.348 \\ 0.419 & 0.342 & 0.386 & -0.671 & -0.245 & 0.22 \\ -0.466 & -0.00136 & 0.302 & 0.269 & -0.586 & 0.525 \end{pmatrix}$$

Example ; 2019-20 Premier League



	PC1	PC2
Liverpool	-4.70	1.20
Manchester City	-4.38	1.65
Manchester United	-2.01	-1.29
Chelsea	-1.29	1.08
Leicester City	-1.66	0.12
Tottenham Hotspur	-0.91	-0.65
Wolverhampton	-0.82	-1.88
Arsenal	-0.46	-1.56
Sheffield United	-0.18	-1.38
Burnley	0.18	-0.10

END