8.1 Introduction

Principal Component Analysis :

- concerned with explaining the variance-covariance structure of a set of variables through a few linear combinations of these variables, for data reduction and interpretation.
- it is more of a means to an end rather than an end in themselves, because they frequently serve as intermediate steps in much larger investigations.

8.2 Population Principal Components

- The linear combinations represent the selection of a new coordinate system obtained by rotating the original system with $X_1, X_2, \dots, X_p$ as the coordinate axes.
- The new axes represent the directions with maximum variability and provide a simpler and more parsimonious description of the covariance structure.

As we shall see, principal components depend solely on the covariance matrix $\boldsymbol{\Sigma}$ (or the correlation matrix $\boldsymbol{\rho}$) of $X_1, X_2, \dots, X_p$. Their development does not require a multivariate normal assumption. On the other hand, principal components derived for multivariate normal populations have useful interpretations in terms of the constant density ellipsoids. Further, inferences can be made

- **Result 8.1.** Let $\boldsymbol{\Sigma}$ be the covariance matrix associated with the random vector $\mathbf{X}' = [X_1, X_2, \dots, X_p]$. Let $\boldsymbol{\Sigma}$ have the eigenvalue-eigenvector pairs $(\lambda_1, \mathbf{e}_1)$, $(\lambda_2, \mathbf{e}_2), \dots, (\lambda_p, \mathbf{e}_p)$ where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. Then the *ith principal component* is given by

$$Y_i = \mathbf{e}_i'\mathbf{X} = e_{i1}X_1 + e_{i2}X_2 + \cdots + e_{ip}X_p, \qquad i = 1, 2, \dots, p \tag{8-4}$$

With these choices,

$$\begin{aligned} \operatorname{Var}(Y_i) &= \mathbf{e}_i'\boldsymbol{\Sigma}\mathbf{e}_i = \lambda_i \qquad i = 1, 2, \dots, p \\ \operatorname{Cov}(Y_i, Y_k) &= \mathbf{e}_i'\boldsymbol{\Sigma}\mathbf{e}_k = 0 \qquad i \neq k \end{aligned} \tag{8-5}$$

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

- the principal components are uncorrelated and have variances equal to the eigenvalues of Σ

Result 8.2. Let $\mathbf{X}' = [X_1, X_2, \dots, X_p]$ have covariance matrix $\boldsymbol{\Sigma}$, with eigenvalue-eigenvector pairs $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \dots, (\lambda_p, \mathbf{e}_p)$ where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. Let $Y_1 = \mathbf{e}_1'\mathbf{X}, Y_2 = \mathbf{e}_2'\mathbf{X}, \dots, Y_p = \mathbf{e}_p'\mathbf{X}$ be the principal components. Then

$$\sigma_{11} + \sigma_{22} + \cdots + \sigma_{pp} = \sum_{i=1}^{p} \operatorname{Var}(X_i) = \lambda_1 + \lambda_2 + \cdots + \lambda_p = \sum_{i=1}^{p} \operatorname{Var}(Y_i)$$

$$\left(\begin{array}{c}\text{Proportion of total} \\ \text{population variance} \\ \text{due to } k\text{th principal} \\ \text{component}\end{array}\right) = \frac{\lambda_k}{\lambda_1 + \lambda_2 + \cdots + \lambda_p} \qquad k = 1, 2, \dots, p \tag{8-7}$$

Result 8.3. If $Y_1 = \mathbf{e}_1'\mathbf{X}$, $Y_2 = \mathbf{e}_2'\mathbf{X}, \ldots, Y_p = \mathbf{e}_p'\mathbf{X}$ are the principal components obtained from the covariance matrix $\mathbf{\Sigma}$, then

$$\rho_{Y_i, X_k} = \frac{e_{ik}\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}} \qquad i, k = 1, 2, \ldots, p \tag{8-8}$$

are the correlation coefficients between the components Y_i and the variables X_k. Here $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \ldots, (\lambda_p, \mathbf{e}_p)$ are the eigenvalue–eigenvector pairs for $\mathbf{\Sigma}$.

* Check Example 8.1 on pdf 455

Principal Components Obtained from Standardized Variables

Result 8.4. The ith principal component of the standardized variables $\mathbf{Z}' = [Z_1, Z_2, \ldots, Z_p]$ with $\text{Cov}(\mathbf{Z}) = \boldsymbol{\rho}$, is given by

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

Moreover,

$$\sum_{i=1}^{p} \text{Var}(Y_i) = \sum_{i=1}^{p} \text{Var}(Z_i) = p \tag{8-11}$$

and

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

In this case, $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \ldots, (\lambda_p, \mathbf{e}_p)$ are the eigenvalue–eigenvector pairs for $\boldsymbol{\rho}$, with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$.

Proof. Result 8.4 follows from Results 8.1, 8.2, and 8.3, with $Z_1, Z_2, \ldots, Z_p$ in place of $X_1, X_2, \ldots, X_p$ and $\boldsymbol{\rho}$ in place of $\boldsymbol{\Sigma}$. ■

Principal Components for Covariance Matrices

- Let Σ be a variance-covariance matrix that all of off-diagonal elements are 0. Then (σ_{ii}, e_i) is the eigenvalue-eigenvector pairs, and the set of principal components is just the original set of uncorrelated random variables.

correlated random variables.

For a covariance matrix with the pattern of (8-13), nothing is gained by extracting the principal components. From another point of view, if $\mathbf{X}$ is distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the contours of constant density are ellipsoids whose axes already lie in the directions of maximum variation. Consequently, there is no need to rotate the coordinate system.

8.3 ## Summarizing Sample Variation by Principal Components

- The sample principal components are defined as those linear combinations which have maximum sample variance

If $\mathbf{S} = \{s_{ik}\}$ is the $p \times p$ sample covariance matrix with eigenvalue-eigenvector pairs $(\hat{\lambda}_1, \hat{\mathbf{e}}_1), (\hat{\lambda}_2, \hat{\mathbf{e}}_2), \ldots, (\hat{\lambda}_p, \hat{\mathbf{e}}_p)$, the ith sample principal component is given by

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

where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_p \geq 0$ and $\mathbf{x}$ is any observation on the variables $X_1, X_2, \ldots, X_p$. Also,

$$\text{Sample variance}(\hat{y}_k) = \hat{\lambda}_k, \quad k = 1, 2, \ldots, p$$
$$\text{Sample covariance}(\hat{y}_i, \hat{y}_k) = 0, \quad i \neq k$$

In addition, (8-20)

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

and

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

The observations $\mathbf{x}_j$ are often "centered" by subtracting $\bar{\mathbf{x}}$. This has no effect on the sample covariance matrix $\mathbf{S}$ and gives the ith principal component

$$\hat{y}_i = \hat{\mathbf{e}}_i'(\mathbf{x} - \bar{\mathbf{x}}), \qquad i = 1, 2, \ldots, p \tag{8-21}$$

for any observation vector $\mathbf{x}$. If we consider the *values* of the ith component

$$\hat{y}_{ji} = \hat{\mathbf{e}}_i'(\mathbf{x}_j - \bar{\mathbf{x}}), \qquad j = 1, 2, \ldots, n \tag{8-22}$$

generated by substituting each observation $\mathbf{x}_j$ for the arbitrary $\mathbf{x}$ in (8-21), then

$$\bar{\hat{y}}_i = \frac{1}{n}\sum_{j=1}^{n} \hat{\mathbf{e}}_i'(\mathbf{x}_j - \bar{\mathbf{x}}) = \frac{1}{n}\hat{\mathbf{e}}_i'\left(\sum_{j=1}^{n}(\mathbf{x}_j - \bar{\mathbf{x}})\right) = \frac{1}{n}\hat{\mathbf{e}}_i'\mathbf{0} = 0 \tag{8-23}$$

That is, the sample mean of each principal component is zero. The sample variances are still given by the $\hat{\lambda}_i$'s, as in (8-20).

The Number of Principal Components

- Things to consider to select a number of principal components
  1. the amount of total sample variance explained
  2. the relative sizes of the eigenvalues
  3. the subject-matter interpretations of the components

Interpretation of the Sample Principal Components

- the sample principal component, $|\hat{y}_i| = |\hat{e}_i'(x-\bar{x})|$, gives the length of the projections of $x-\bar{x}$ in the directions of the axes $\hat{e}_i$.

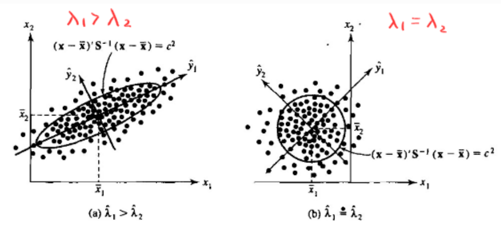

Figure 8.4 Sample principal components and ellipses of constant distance.

- When the contours of constant distance are nearly circular or, equivalently, when the eigenvalues of S are nearly equal, the sample variation is homogeneous in all directions

Standardizing the Sample Principal Components

☆ 위에 population PC랑 내용 비슷비슷한거 같은데

8.4 Graphing the Principal Components

쓸데 없어보이고

8.5 Large Sample Inferences

- The eigenvectors determine the directions of maximum variability, and the eigenvalues specify the variances.
- It is important to note that because of sampling variation, the eigenvalues and eigenvectors will differ from their underlying population counterparts.

Large Sample Properties of $\hat{\lambda}_i$ and $\hat{e}_i$

~ 8.6 까지 Out of 범위

Supplement 8A The Geometry of the Sample Principal Component Approximation