연세대학교 통계 데이터 사이언스 학회 ESC 23-2 FALL WEEK3

Factor Analysis

[ESC 정규세션 학술부] 임승현 전인태

1. Introduction

Introduction

Factor Analysis

· 변수들 간의 covariance를 고려하여 데이터를 나타내는 변수들을 더 적은 수의 요인(factor) 으로 제시해주는 분석 방법.

→ 변수들 간의 correlation을 고려하여 변수들을 그룹화해준다. 이 때 수많은 변수들이 있을 때, 그들 사이에 어떤 관계가 있어서 (특정 변수가 변하면 다른 변수도 변하는 등), 그 관계를 공통적으로 설명하는 숨겨진 요인이 있다는 가정 하에 그 요인(factor)을 찾게 된다.

· Factor Analysis의 목적

1) 새 변수의 생성 효과 및 전처리 기능

→ 다른 분석 방법에 데이터들을 입력하기에 앞서, 본래의 변수보다 더 적절한 변수들을 생성, 더 좋은 분석 결과나 모델을 얻을 수 있음.

2) 데이터 축소

→ 데이터 축소를 통해 분석 결과의 해석이 쉬워짐. 다중공선성 문제 해결할 수 있음.

Introduction

Example

· 각각의 학생들이 (수학, 과학, 영어, 독일어, 체육, 미술)의 시험 성적을 가지고 있다. 이를 $(X_1, \cdots, X_6)$ 라고 하자. 이 때 학생들의 성적을 더 적은 수의 변수로 나타내는 방법은 다음과 같다.

예를 들어. (X_1, X_2), 즉 (수학, 과학) 성적은 "수리 능력" 이라는 factor로 묶을 수 있을 것이다.
마찬가지로 (X_3, X_4)인 (영어, 독일어) 성적은 "외국어 능력"이라는 factor로,
(X_5, X_6)인 (체육, 미술) 성적은 "예체능 능력"이라는 factor로 묶일 수 있다.

· 다음 오른쪽 표와 같이 변수들이 요약될 수 있다.
F_1은 (X_1, X_2)에 의해, F_2는 (X_3, X_4)에 의해, F_3는 (X_5, X_6)에 의해 설명된다.

	X_1	X_2	X_3	X_4	X_5	X_6
F_1	a_1X_1	a_2X_2				
F_2			a_3X_3	a_4X_4		
F_3					a_5X_5	a_6X_6

Introduction

FA vs PCA

· 공통점

1) 데이터의 차원을 축소하는 목적을 가지고 있다.
2) 변수 간의 상관 관계, 공분산을 바탕으로 하는 분석이다.

· 차이점

1) 생성되는 변수의 의미

→ FA는 생성된 factor마다 각각의 의미를 부여할 수 있다. (수리 능력, 외국어 능력)
PCA는 주성분을 통해 변수들을 잘 설명할 수 있는 임의의 변수를 만들기 때문에 생성된 변수에 의미를 부여하는 것이 어렵다

2) 생성되는 변수의 개수

→ FA와 PCA 모두 특정 기준에 따라 (Ex) 생성된 변수가 기존의 변수를 얼마나 설명하는지) 따라 생성되는 변수의 개수가 보통은 정해져 있다. 하지만 FA의 경우, 필요에 따라 prior knowledge나 데이터 분석의 목적 등을 반영하여 common factor의 개수를 추정할 수 있다.

2. The Orthogonal Factor Model

The Orthogonal Factor Model

The Orthogonal Factor Model

$$\underset{(p\times 1)}{\mathbf{X}-\boldsymbol{\mu}} = \underset{(p\times m)}{\mathbf{L}}\underset{(m\times 1)}{\mathbf{F}} + \underset{(p\times 1)}{\boldsymbol{\varepsilon}}$$

$$\begin{aligned}
X_1 - \mu_1 &= \ell_{11}F_1 + \ell_{12}F_2 + \cdots + \ell_{1m}F_m + \varepsilon_1 \\
X_2 - \mu_2 &= \ell_{21}F_1 + \ell_{22}F_2 + \cdots + \ell_{2m}F_m + \varepsilon_2 \\
&\vdots \\
X_p - \mu_p &= \ell_{p1}F_1 + \ell_{p2}F_2 + \cdots + \ell_{pm}F_m + \varepsilon_p
\end{aligned}$$

· $F_1, F_2, \dots, F_m$과 $\epsilon_1, \epsilon_2 \dots, \epsilon_p$는 unobservable

→ random vector F, ϵ에 대한 추가적인 가정 필요!

· 관측 가능한 p차원의 vector X가 mean μ, covariance Σ 를 가짐.

· X는 common factors로 불리는 $F_1, F_2, \dots, F_m$ 에 대해 linearly dependent.

· $\epsilon_1, \epsilon_2 \dots, \epsilon_p$는 error

· Coefficient l_{ij}는 loading, 행렬 L은 matrix of factor loadings

The Orthogonal Factor Model

The Orthogonal Factor Model

$$\underset{(p\times 1)}{\mathbf{X}-\boldsymbol{\mu}} = \underset{(p\times m)}{\mathbf{L}}\,\underset{(m\times 1)}{\mathbf{F}} + \underset{(p\times 1)}{\boldsymbol{\varepsilon}}$$

· $F_1, F_2, \dots, F_m$과 $\epsilon_1, \epsilon_2 \dots, \epsilon_p$는 unobservable

→ random vector F, ϵ에 대한 추가적인 가정 필요!

$$E(\mathbf{F}) = \underset{(m\times 1)}{\mathbf{0}}, \qquad \operatorname{Cov}(\mathbf{F}) = E[\mathbf{F}\mathbf{F}'] = \underset{(m\times m)}{\mathbf{I}}$$

$$E(\boldsymbol{\varepsilon}) = \underset{(p\times 1)}{\mathbf{0}}, \qquad \operatorname{Cov}(\boldsymbol{\varepsilon}) = E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \underset{(p\times p)}{\boldsymbol{\Psi}} = \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_p \end{bmatrix}$$

$$\operatorname{Cov}(\boldsymbol{\varepsilon}, \mathbf{F}) = E(\boldsymbol{\varepsilon}\mathbf{F}') = \underset{(p\times m)}{\mathbf{0}}$$

random vector F, ϵ는 independent

The Orthogonal Factor Model

Orthogonal Factor Model with m Common Factors

$$\underset{(p\times 1)}{\mathbf{X}} = \underset{(p\times 1)}{\boldsymbol{\mu}} + \underset{(p\times m)}{\mathbf{L}}\underset{(m\times 1)}{\mathbf{F}} + \underset{(p\times 1)}{\boldsymbol{\varepsilon}}$$

$$\begin{aligned} \mu_i &= \textit{mean} \text{ of variable } i \\ \varepsilon_i &= i\text{th } \textit{specific factor} \\ F_j &= j\text{th } \textit{common factor} \\ \ell_{ij} &= \textit{loading} \text{ of the } i\text{th variable on the } j\text{th factor} \end{aligned} \qquad (9\text{-}4)$$

The unobservable random vectors $\mathbf{F}$ and $\boldsymbol{\varepsilon}$ satisfy the following conditions:

$\mathbf{F}$ and $\boldsymbol{\varepsilon}$ are independent

$E(\mathbf{F}) = \mathbf{0}$, $\text{Cov}(\mathbf{F}) = \mathbf{I}$

$E(\boldsymbol{\varepsilon}) = \mathbf{0}$, $\text{Cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Psi}$, where $\boldsymbol{\Psi}$ is a diagonal matrix

Figure 1

더 적은 변수들로 X를 요약하고 싶다면, X의 차원인 p보다 더 적은 m개의 factor를 설정해야 함

The Orthogonal Factor Model

The Orthogonal Factor Model

$$\underset{(p\times 1)}{\mathbf{X}-\boldsymbol{\mu}} = \underset{(p\times m)}{\mathbf{L}}\,\underset{(m\times 1)}{\mathbf{F}} + \underset{(p\times 1)}{\boldsymbol{\varepsilon}}$$

$$\begin{aligned}(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})' &= (\mathbf{LF}+\boldsymbol{\varepsilon})(\mathbf{LF}+\boldsymbol{\varepsilon})' \\ &= (\mathbf{LF}+\boldsymbol{\varepsilon})((\mathbf{LF})'+\boldsymbol{\varepsilon}') \\ &= \mathbf{LF}(\mathbf{LF})' + \boldsymbol{\varepsilon}(\mathbf{LF})' + \mathbf{LF}\boldsymbol{\varepsilon}' + \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\end{aligned}$$

$$\begin{aligned}\boldsymbol{\Sigma} &= \mathrm{Cov}(\mathbf{X}) = E(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})' \\ &= \mathbf{L}E(\mathbf{FF}')\mathbf{L}' + E(\boldsymbol{\varepsilon}\mathbf{F}')\mathbf{L}' + \mathbf{L}E(\mathbf{F}\boldsymbol{\varepsilon}') + E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') \\ &= \mathbf{LL}' + \boldsymbol{\Psi}\end{aligned}$$

$$(\mathbf{X}-\boldsymbol{\mu})\mathbf{F}' = (\mathbf{LF}+\boldsymbol{\varepsilon})\mathbf{F}' = \mathbf{LF}\,\mathbf{F}' + \boldsymbol{\varepsilon}\mathbf{F}' \qquad \mathrm{Cov}(\boldsymbol{\varepsilon},\mathbf{F}) = E(\boldsymbol{\varepsilon}\mathbf{F}') = \underset{(p\times m)}{\mathbf{0}}$$

$$\mathrm{Cov}(\mathbf{X},\mathbf{F}) = E(\mathbf{X}-\boldsymbol{\mu})\mathbf{F}' = \mathbf{L}E(\mathbf{FF}') + E(\boldsymbol{\varepsilon}\mathbf{F}') = \mathbf{L}$$

The Orthogonal Factor Model

Covariance Structure for the Orthogonal Factor Model

1. $\text{Cov}(\mathbf{X}) = \mathbf{LL}' + \mathbf{\Psi}$

or

$$\text{Var}(X_i) = \ell_{i1}^2 + \cdots + \ell_{im}^2 + \psi_i$$

$$\text{Cov}(X_i, X_k) = \ell_{i1}\ell_{k1} + \cdots + \ell_{im}\ell_{km}$$

2. $\text{Cov}(\mathbf{X}, \mathbf{F}) = \mathbf{L}$

or

$$\text{Cov}(X_i, F_j) = \ell_{ij}$$

$$\begin{aligned} X_1 - \mu_1 &= \ell_{11}F_1 + \ell_{12}F_2 + \cdots + \ell_{1m}F_m + \varepsilon_1 \\ X_2 - \mu_2 &= \ell_{21}F_1 + \ell_{22}F_2 + \cdots + \ell_{2m}F_m + \varepsilon_2 \\ &\vdots \\ X_p - \mu_p &= \ell_{p1}F_1 + \ell_{p2}F_2 + \cdots + \ell_{pm}F_m + \varepsilon_p \end{aligned}$$

Figure 2

The Orthogonal Factor Model

1. $\text{Cov}(\mathbf{X}) = \mathbf{LL}' + \boldsymbol{\Psi}$

   or

$$\text{Var}(X_i) = \ell_{i1}^2 + \cdots + \ell_{im}^2 + \psi_i$$

$$\text{Cov}(X_i, X_k) = \ell_{i1}\ell_{k1} + \cdots + \ell_{im}\ell_{km}$$

· i번째 variable, 즉 X_i의 variance는 i번째에 해당하는 m개의 common factors에 의해 설명됨.

$$\underbrace{\sigma_{ii}}_{\text{Var}(X_i)} = \underbrace{\ell_{i1}^2 + \ell_{i2}^2 + \cdots + \ell_{im}^2}_{\text{communality}} + \underbrace{\psi_i}_{\text{specific variance}}$$

$$h_i^2 = \ell_{i1}^2 + \ell_{i2}^2 + \cdots + \ell_{im}^2$$ → ith communality

$$\sigma_{ii} = h_i^2 + \psi_i, \qquad i = 1, 2, \ldots, p$$

The Orthogonal Factor Model

Example $\boldsymbol{\Sigma} = \mathbf{LL}' + \boldsymbol{\Psi}$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 19 & 30 & 2 & 12 \\ 30 & 57 & 5 & 23 \\ 2 & 5 & 38 & 47 \\ 12 & 23 & 47 & 68 \end{bmatrix} = \begin{bmatrix} 4 & 1 \\ 7 & 2 \\ -1 & 6 \\ 1 & 8 \end{bmatrix} \begin{bmatrix} 4 & 7 & -1 & 1 \\ 1 & 2 & 6 & 8 \end{bmatrix} + \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

Figure 3

$$\mathbf{L} = \begin{bmatrix} \ell_{11} & \ell_{12} \\ \ell_{21} & \ell_{22} \\ \ell_{31} & \ell_{32} \\ \ell_{41} & \ell_{42} \end{bmatrix} = \begin{bmatrix} 4 & 1 \\ 7 & 2 \\ -1 & 6 \\ 1 & 8 \end{bmatrix} \qquad \boldsymbol{\Psi} = \begin{bmatrix} \psi_1 & 0 & 0 & 0 \\ 0 & \psi_2 & 0 & 0 \\ 0 & 0 & \psi_3 & 0 \\ 0 & 0 & 0 & \psi_4 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

$$\sigma_{11} = (\ell_{11}^2 + \ell_{12}^2) + \psi_1 = h_1^2 + \psi_1$$

$$\underbrace{19}_{\text{variance}} = \underbrace{4^2 + 1^2}_{\text{communality}} + \underbrace{2}_{\text{specific variance}} = 17 + 2$$

The Orthogonal Factor Model

The Orthogonal Factor Model

$$\underset{(p\times 1)}{\mathbf{X} - \boldsymbol{\mu}} = \underset{(p\times m)}{\mathbf{L}} \; \underset{(m\times 1)}{\mathbf{F}} + \underset{(p\times 1)}{\boldsymbol{\varepsilon}}$$

$$\begin{aligned}\boldsymbol{\Sigma} &= \text{Cov}(\mathbf{X}) = E(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})' \\ &= \mathbf{L}E(\mathbf{F}\mathbf{F}')\mathbf{L}' + E(\boldsymbol{\varepsilon}\mathbf{F}')\mathbf{L}' + \mathbf{L}E(\mathbf{F}\boldsymbol{\varepsilon}') + E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') \\ &= \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}\end{aligned}$$

· Factor model은 Σ, 즉 총 $p(p+1)/2$개의 variance와 covariance들이 pm개의 loadings l_{ij}와 p개의 특정한 variance ψ 로 reproduce될 수 있다는 가정을 하고 있음.

· $p > m$인 경우, 총 p개의 변수들이 m개의 변수들로 요약될 수 있으므로 이는 데이터를 축소할 수 있음을 의미

Ex) X가 $p = 12$개의 변수들로 이루어져 있고 $m = 2$개의 factor로 요약하는 것이 적절한 경우

→Σ는 $\frac{12(13)}{2} = 78$개의 element로 구성되지만 이를 $mp + p = 12(2) + 12 = 36$개의 parameter로 요약할 수 있음.

· 하지만 X를 구성하는 p개의 변수에 비해 m이 너무 작은 경우, $\underset{(p\times 1)}{\mathbf{X} - \boldsymbol{\mu}} = \underset{(p\times m)}{\mathbf{L}} \; \underset{(m\times 1)}{\mathbf{F}} + \underset{(p\times 1)}{\boldsymbol{\varepsilon}}$ 로 표현이 불가능할 수도 있음.(Ex 9.2)

ESC

The Orthogonal Factor Model

The Idea of Factor Rotation

$$\mathbf{T}\mathbf{T}' = \mathbf{T}'\mathbf{T} = \mathbf{I}$$

T는 $m \times m$인orthogonal matrix

$$\mathbf{X} - \boldsymbol{\mu} = \mathbf{L}\mathbf{F} + \boldsymbol{\varepsilon} = \mathbf{L}\mathbf{T}\mathbf{T}'\mathbf{F} + \boldsymbol{\varepsilon} = \mathbf{L}^*\mathbf{F}^* + \boldsymbol{\varepsilon}$$

즉, T는 orthogonal operator 역할을 하여 기존의 F와 L을 rotation시킴.

$$\mathbf{L}^* = \mathbf{L}\mathbf{T} \quad \text{and} \quad \mathbf{F}^* = \mathbf{T}'\mathbf{F}$$

$$E(\mathbf{F}^*) = \mathbf{T}'E(\mathbf{F}) = \mathbf{0}$$

$$\text{Cov}(\mathbf{F}^*) = \mathbf{T}'\text{Cov}(\mathbf{F})\mathbf{T} = \mathbf{T}'\mathbf{T} = \underset{(m\times m)}{\mathbf{I}}$$

F와 $F^* = T'F$는 동일한 statistical property를 갖는다.

$$\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi} = \mathbf{L}\mathbf{T}\mathbf{T}'\mathbf{L}' + \boldsymbol{\Psi} = (\mathbf{L}^*)(\mathbf{L}^*)' + \boldsymbol{\Psi}$$

· 최종적으로 $\mathbf{L}^* = \mathbf{L}\mathbf{T} \quad \text{and} \quad \mathbf{F}^* = \mathbf{T}'\mathbf{F}$은 동일한 covariance matrix Σ를 생성한다.

→ orthogonal operator 역할을 하는 T를 여러 번 바꿔가며 L과 F에 적용해보면, F에 대한 해석을 좀 더 용이하게 할 수 없을까?

→ "Factor Rotation"

3. Methods of Estimation

Methods of Estimation

Transformation

$$\mathbf{x}_j - \bar{\mathbf{x}} = \begin{bmatrix} x_{j1} \\ x_{j2} \\ \vdots \\ x_{jp} \end{bmatrix} - \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix} = \begin{bmatrix} x_{j1} - \bar{x}_1 \\ x_{j2} - \bar{x}_2 \\ \vdots \\ x_{jp} - \bar{x}_p \end{bmatrix} \qquad j = 1, 2, \dots, n$$

· 각각의 observation들에 sample mean을 빼주는 방법.

· 변환된 데이터들의 covariance matrix S는 기존의 Σ와 동일하다.

$$\mathbf{z}_j = \begin{bmatrix} \dfrac{(x_{j1} - \bar{x}_1)}{\sqrt{s_{11}}} \\ \dfrac{(x_{j2} - \bar{x}_2)}{\sqrt{s_{22}}} \\ \vdots \\ \dfrac{(x_{jp} - \bar{x}_p)}{\sqrt{s_{pp}}} \end{bmatrix} \qquad j = 1, 2, \dots, n$$

· 각각의 변수들을 standardized 해주는 방법

· transformation된 변수들은 sample correlation matrix R을 가지게 된다.

Methods of Estimation

Methods of Estimation

· Factor loading인 l_{ij}와 specific variance ψ_i를 estimate 하는 과정

· Principal component method (Principal factor method), maximum likelihood method가 있다.

Principal Component Method

· Spectral decomposition을 이용하여 covariance Σ를 추정하는 방법

$$\boldsymbol{\Sigma} \doteq \left[\sqrt{\lambda_1}\,\mathbf{e}_1 \;\vdots\; \sqrt{\lambda_2}\,\mathbf{e}_2 \;\vdots\; \cdots \;\vdots\; \sqrt{\lambda_m}\,\mathbf{e}_m \right] \begin{bmatrix} \sqrt{\lambda_1}\,\mathbf{e}_1' \\ \hdashline \sqrt{\lambda_2}\,\mathbf{e}_2' \\ \hdashline \vdots \\ \hdashline \sqrt{\lambda_m}\,\mathbf{e}_m' \end{bmatrix} = \underset{(p\times m)}{\mathbf{L}} \; \underset{(m\times p)}{\mathbf{L}'}$$

· $\lambda_1 \geq \cdots \geq \lambda_m \geq 0$을 만족. (PCA의 주성분을 구하는 과정과 같다. 주성분을 이용하여 Σ를 추정.

· j번째 factor에 대한 factor loading들은 j번째 주성분에 대한 coefficient가 된다.

Methods of Estimation

Principal Component Method

$$\boldsymbol{\Sigma} \doteq \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}$$

$$= \left[\sqrt{\lambda_1}\,\mathbf{e}_1 \;\vdots\; \sqrt{\lambda_2}\,\mathbf{e}_2 \;\vdots\; \cdots \;\vdots\; \sqrt{\lambda_m}\,\mathbf{e}_m\right] \begin{bmatrix} \sqrt{\lambda_1}\,\mathbf{e}_1' \\ \hdashline \sqrt{\lambda_2}\,\mathbf{e}_2' \\ \hdashline \vdots \\ \hdashline \sqrt{\lambda_m}\,\mathbf{e}_m' \end{bmatrix} + \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_p \end{bmatrix}$$

$$\psi_i = \sigma_{ii} - \sum_{j=1}^{m} \ell_{ij}^2 \text{ for } i = 1, 2, \ldots, p.$$

· $p = m$인 경우가 아닌 이상, Σ와 LL'은 같아질 수 없고, specific factor를 더하여 등식을 성립하게 해준다. (Error term에 대한 variance)

Methods of Estimation

Principal Component Solution of the Factor Model

The principal component factor analysis of the sample covariance matrix $\mathbf{S}$ is specified in terms of its eigenvalue–eigenvector pairs $(\hat{\lambda}_1, \hat{\mathbf{e}}_1)$, $(\hat{\lambda}_2, \hat{\mathbf{e}}_2), \ldots,$ $(\hat{\lambda}_p, \hat{\mathbf{e}}_p)$, where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_p$. Let $m < p$ be the number of common factors. Then the matrix of estimated factor loadings $\{\tilde{\ell}_{ij}\}$ is given by

$$\tilde{\mathbf{L}} = \left[\sqrt{\hat{\lambda}_1}\, \hat{\mathbf{e}}_1 \;\vdots\; \sqrt{\hat{\lambda}_2}\, \hat{\mathbf{e}}_2 \;\vdots\; \cdots \;\vdots\; \sqrt{\hat{\lambda}_m}\, \hat{\mathbf{e}}_m \right] \tag{9-15}$$

The estimated specific variances are provided by the diagonal elements of the matrix $\mathbf{S} - \tilde{\mathbf{L}}\tilde{\mathbf{L}}'$, so

$$\tilde{\mathbf{\Psi}} = \begin{bmatrix} \tilde{\psi}_1 & 0 & \cdots & 0 \\ 0 & \tilde{\psi}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \tilde{\psi}_p \end{bmatrix} \quad \text{with} \quad \tilde{\psi}_i = s_{ii} - \sum_{j=1}^{m} \tilde{\ell}_{ij}^2 \tag{9-16}$$

Communalities are estimated as

$$\tilde{h}_i^2 = \tilde{\ell}_{i1}^2 + \tilde{\ell}_{i2}^2 + \cdots + \tilde{\ell}_{im}^2 \tag{9-17}$$

Figure 3

Methods of Estimation

Principal Component Method

· 어떻게 적절한 m을 선택? → Residual $\mathbf{S} - (\tilde{\mathbf{L}}\tilde{\mathbf{L}}' + \tilde{\mathbf{\Psi}})$ 을 고려하자

$$(\mathbf{S} - (\tilde{\mathbf{L}}\tilde{\mathbf{L}}' + \tilde{\mathbf{\Psi}})) \leq \hat{\lambda}_{m+1}^2 + \cdots + \hat{\lambda}_p^2$$

· total sample variance $\left(s_{11} + s_{22} + \cdots + s_{pp} = tr(S)\right)$

$$\begin{pmatrix}\text{Proportion of total} \\ \text{sample variance} \\ \text{due to } j\text{th factor}\end{pmatrix} = \begin{cases} \dfrac{\hat{\lambda}_j}{s_{11} + s_{22} + \cdots + s_{pp}} & \text{for a factor analysis of } \mathbf{S} \\ \dfrac{\hat{\lambda}_j}{p} & \text{for a factor analysis of } \mathbf{R} \end{cases}$$

$$\tilde{\ell}_{11}^2 + \tilde{\ell}_{21}^2 + \cdots + \tilde{\ell}_{p1}^2 = \left(\sqrt{\hat{\lambda}_1}\,\hat{\mathbf{e}}_1\right)'\left(\sqrt{\hat{\lambda}_1}\,\hat{\mathbf{e}}_1\right) = \hat{\lambda}_1$$

Recall) 1주차 세션 내용

THEOREM 9.1 For a given $X \sim (\mu, \Sigma)$ let $Y = \Gamma^\top(X - \mu)$

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

$$Var(Y_j) = \lambda_j, \quad j = 1, \ldots, p$$

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

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

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

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

Methods of Estimation

Maximum Likelihood Method

· Common factor F와 specific factor ϵ이 normally distributed 하다는 가정

F_i, ϵ_i가 jointly normal 하다면, $\mathbf{X}_j - \boldsymbol{\mu} = \mathbf{L}\mathbf{F}_j + \boldsymbol{\varepsilon}_j$ 은 normal을 따르게 된다.

→ F, ϵ에 대한 maximum likelihood estimate를 구할 수 있다

$$
\begin{aligned}
L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= (2\pi)^{-\frac{np}{2}} |\boldsymbol{\Sigma}|^{-\frac{n}{2}} e^{-\left(\frac{1}{2}\right)\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n} (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})' + n(\bar{\mathbf{x}} - \boldsymbol{\mu})(\bar{\mathbf{x}} - \boldsymbol{\mu})'\right)\right]} \\
&= (2\pi)^{\frac{-(n-1)p}{2}} |\boldsymbol{\Sigma}|^{-\frac{(n-1)}{2}} e^{-\left(\frac{1}{2}\right)\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n} (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})'\right)\right]} \\
&\quad \times (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} e^{-\left(\frac{n}{2}\right)(\bar{\mathbf{x}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu})}
\end{aligned}
$$

n은 관측치의 개수, p는 X의 component의 개수

$\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}$

Subject to $\mathbf{L}'\boldsymbol{\Psi}^{-1}\mathbf{L} = \boldsymbol{\Delta}$ (diagonal matrix)

→Uniqueness condition을 설정해야 L이 well defined

· 위의 likelihood를 maximization하면 maximum likelihood estimates $\hat{L}, \hat{\Psi}, \hat{\mu} = \bar{x}$를 얻을 수 있다.

Methods of Estimation

Maximum Likelihood Method

· Principal component method와 마찬가지로, total variance에 대한 j번째 factor의 variance비율을 구할 수 있음.

$$\begin{pmatrix} \text{Proportion of total sample} \\ \text{variance due to } j\text{th factor} \end{pmatrix} = \frac{\hat{\ell}_{1j}^2 + \hat{\ell}_{2j}^2 + \cdots + \hat{\ell}_{pj}^2}{s_{11} + s_{22} + \cdots + s_{pp}}$$

$$\begin{pmatrix} \text{Proportion of total (standardized)} \\ \text{sample variance due to } j\text{th factor} \end{pmatrix} = \frac{\hat{\ell}_{1j}^2 + \hat{\ell}_{2j}^2 + \cdots + \hat{\ell}_{pj}^2}{p}$$

· j번째 factor에 대한 상대적인 중요성을 알 수 있다.

Methods of Estimation

Example 9.3

$$
\begin{array}{lc}
\textit{Attribute (Variable)} & \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\
\begin{array}{l} \text{Taste} \\ \text{Good buy for money} \\ \text{Flavor} \\ \text{Suitable for snack} \\ \text{Provides lots of energy} \end{array}
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}
& \begin{bmatrix} 1.00 & .02 & .96 & .42 & .01 \\ .02 & 1.00 & .13 & .71 & .85 \\ .96 & .13 & 1.00 & .50 & .11 \\ .42 & .71 & .50 & 1.00 & .79 \\ .01 & .85 & .11 & .79 & 1.00 \end{bmatrix}
\end{array}
$$

Figure 4

· 새로운 판매 상품에 대한 고객들의 평가. 총 5개의 component에 대해 데이터를 수집하였고, 수집한 데이터에 대한 correlation matrix는 왼쪽과 같다

· correlation matrix로부터 (1, 3), (2, 5)는 그룹을 이룬다. 이 때 variable 4 또한 group (2, 5)에 더 가깝다는 것을 알 수 있다.

$$
\tilde{\mathbf{L}}\tilde{\mathbf{L}}' + \tilde{\boldsymbol{\Psi}} = \begin{bmatrix} .56 & .82 \\ .78 & -.53 \\ .65 & .75 \\ .94 & -.10 \\ .80 & -.54 \end{bmatrix} \begin{bmatrix} .56 & .78 & .65 & .94 & .80 \\ .82 & -.53 & .75 & -.10 & -.54 \end{bmatrix}
$$

$$
+ \begin{bmatrix} .02 & 0 & 0 & 0 & 0 \\ 0 & .12 & 0 & 0 & 0 \\ 0 & 0 & .02 & 0 & 0 \\ 0 & 0 & 0 & .11 & 0 \\ 0 & 0 & 0 & 0 & .07 \end{bmatrix} = \begin{bmatrix} 1.00 & .01 & .97 & .44 & .00 \\ & 1.00 & .11 & .79 & .91 \\ & & 1.00 & .53 & .11 \\ & & & 1.00 & .81 \\ & & & & 1.00 \end{bmatrix}
$$

Methods of Estimation

Table 9.1				
Variable	Estimated factor loadings $\tilde{\ell}_{ij} = \sqrt{\hat{\lambda}_i}\,\hat{e}_{ij}$ F_1	F_2	Communalities $\tilde{h}_i^2$	Specific variances $\tilde{\psi}_i = 1 - \tilde{h}_i^2$
1. Taste	.56	.82	.98	.02
2. Good buy for money	.78	−.53	.88	.12
3. Flavor	.65	.75	.98	.02
4. Suitable for snack	.94	−.10	.89	.11
5. Provides lots of energy	.80	−.54	.93	.07
Eigenvalues	2.85	1.81		
Cumulative proportion of total (standardized) sample variance	.571	.932		

Figure 4

$$\tilde{\mathbf{L}}\tilde{\mathbf{L}}' + \tilde{\mathbf{\Psi}} = \begin{bmatrix} .56 & .82 \\ .78 & -.53 \\ .65 & .75 \\ .94 & -.10 \\ .80 & -.54 \end{bmatrix} \begin{bmatrix} .56 & .78 & .65 & .94 & .80 \\ .82 & -.53 & .75 & -.10 & -.54 \end{bmatrix}$$

$$+ \begin{bmatrix} .02 & 0 & 0 & 0 & 0 \\ 0 & .12 & 0 & 0 & 0 \\ 0 & 0 & .02 & 0 & 0 \\ 0 & 0 & 0 & .11 & 0 \\ 0 & 0 & 0 & 0 & .07 \end{bmatrix} = \begin{bmatrix} 1.00 & .01 & .97 & .44 & .00 \\ & 1.00 & .11 & .79 & .91 \\ & & 1.00 & .53 & .11 \\ & & & 1.00 & .81 \\ & & & & 1.00 \end{bmatrix}$$

Methods of Estimation

Table 9.1

Variable	Estimated factor loadings $\tilde{\ell}_{ij} = \sqrt{\hat{\lambda}_i}\,\hat{e}_{ij}$ F_1	F_2	Communalities $\tilde{h}_i^2$	Specific variances $\tilde{\psi}_i = 1 - \tilde{h}_i^2$
1. Taste	.56	.82	.98	.02
2. Good buy for money	.78	−.53	.88	.12
3. Flavor	.65	.75	.98	.02
4. Suitable for snack	.94	−.10	.89	.11
5. Provides lots of energy	.80	−.54	.93	.07
Eigenvalues	2.85	1.81		
Cumulative proportion of total (standardized) sample variance	.571	.932		

Figure 4

· 두 개의 factor F_1, F_2로 total (standardized) sample variance의 $\frac{(\widehat{\lambda_1}+\widehat{\lambda_2})}{5} = \frac{2.85+1.81}{5} = 0.93$ 정도가 설명된다.

· F_1은 "상품의 전반적인 평가"를 설명한다고 말할 수 있다.

· F_2는 "Taste와 Flavor에 대한 평가"를 설명한다고 할 수 있다.

· Communality인 (0.98, 0.88, 0.98, 0.89, 0.93)은 모두 1에 가까운 숫자를 보이고 있다. 이는 두 개의 factor가 5개의 변수에 대한 sample variance를 높은 비율로 잘 설명하고 있다는 것을 의미한다.

Methods of Estimation

A Large Sample Test for the Number of Common Factors

· m개의 factor을 이용하여 factor model을 구성하였을 때, factor model이 covariance matrix를 잘 반영하여 구성되었는가?

· Likelihood ratio test 이용!

$$H_0: \underset{(p\times p)}{\boldsymbol{\Sigma}} = \underset{(p\times m)}{\mathbf{L}}\ \underset{(m\times p)}{\mathbf{L}'} + \underset{(p\times p)}{\boldsymbol{\Psi}}$$

$$-2\ln\Lambda = -2\ln\left[\frac{\text{maximized likelihood under } H_0}{\text{maximized likelihood}}\right]$$

$$\begin{aligned} L(\boldsymbol{\mu},\boldsymbol{\Sigma}) &= (2\pi)^{-\frac{np}{2}}|\boldsymbol{\Sigma}|^{-\frac{n}{2}}e^{-\left(\frac{1}{2}\right)\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\mathbf{x}_j-\bar{\mathbf{x}})(\mathbf{x}_j-\bar{\mathbf{x}})'+n(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})'\right)\right]} \\ &= (2\pi)^{-\frac{(n-1)p}{2}}|\boldsymbol{\Sigma}|^{-\frac{(n-1)}{2}}e^{-\left(\frac{1}{2}\right)\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\mathbf{x}_j-\bar{\mathbf{x}})(\mathbf{x}_j-\bar{\mathbf{x}})'\right)\right]} \\ &\quad\times(2\pi)^{-\frac{p}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}e^{-\left(\frac{n}{2}\right)(\bar{\mathbf{x}}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu})} \end{aligned}$$

Methods of Estimation

$$H_0: \underset{(p\times p)}{\boldsymbol{\Sigma}} = \underset{(p\times m)}{\mathbf{L}} \underset{(m\times p)}{\mathbf{L}'} + \underset{(p\times p)}{\boldsymbol{\Psi}} \qquad -2\ln\Lambda = -2\ln\left[\frac{\text{maximized likelihood under } H_0}{\text{maximized likelihood}}\right]$$

$$L(\boldsymbol{\mu},\boldsymbol{\Sigma}) = (2\pi)^{-\frac{np}{2}}|\boldsymbol{\Sigma}|^{-\frac{n}{2}}e^{-\left(\frac{1}{2}\right)\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\mathbf{x}_j-\bar{\mathbf{x}})(\mathbf{x}_j-\bar{\mathbf{x}})'+n(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})'\right)\right]}$$

1) Maximized likelihood under H_0 => likelihood에 $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}$ 를 대입하고 constant term은 미리 정리

$$|\hat{\boldsymbol{\Sigma}}|^{-n/2}\exp\left(-\tfrac{1}{2}\operatorname{tr}\left[\hat{\boldsymbol{\Sigma}}^{-1}\left(\sum_{j=1}^{n}(\mathbf{x}_j-\bar{\mathbf{x}})(\mathbf{x}_j-\bar{\mathbf{x}})'\right)\right]\right) = |\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}}|^{-n/2}\exp\left(-\tfrac{1}{2}n\operatorname{tr}\left[(\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}})^{-1}\mathbf{S}_n\right]\right)$$

2) Maximized likelihood => likelihood에 mle $\hat{L}, \hat{\Psi}, \hat{\mu} = \bar{x}$ 를 대입하고 constant term은 미리 정리

$$|\mathbf{S}_n|^{-n/2}e^{-np/2} \qquad \text{where} \qquad (n-1)/n)\mathbf{S} = \mathbf{S}_n$$

3) 자유도 계산

$$\begin{aligned} v - v_0 &= \tfrac{1}{2}p(p+1) - \left[p(m+1) - \tfrac{1}{2}m(m-1)\right] \\ &= \tfrac{1}{2}\left[(p-m)^2 - p - m\right] \end{aligned}$$

Methods of Estimation

$$-2\ln\Lambda = -2\ln\left[\frac{\text{maximized likelihood under } H_0}{\text{maximized likelihood}}\right]$$

$$= -2\ln\left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\mathbf{S}_n|}\right)^{-n/2} + n\left[\operatorname{tr}(\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S}_n) - p\right]$$

$$= n\ln\left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\mathbf{S}_n|}\right)$$

By supplement 9A

$$\frac{|\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}}|^{-n/2}\exp\left(-\tfrac{1}{2}n\operatorname{tr}[(\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}})^{-1}\mathbf{S}_n]\right)}{|\mathbf{S}_n|^{-n/2}e^{-np/2}}$$

Bartlett's correction을 적용, 아래를 만족하면 유의 수준 α에서 reject H_0

$$(n - 1 - (2p + 4m + 5)/6)\ln\frac{|\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}}|}{|\mathbf{S}_n|} > \chi^2_{[(p-m)^2-p-m]/2}(\alpha)$$

4.Factor Rotation

Factor Rotation

Factor Rotation

우리가 factor analysis에 기대하는 이상적인 상황은 각각의 변수가 하나의 common factor에 대해서만 큰 loading 값을 가지고, 다른 나머지 factor 들에 대해서는 작은 loading 값을 가지는 것이다.

그런데 어떤 factor analysis의 결과, 오른쪽 표와 같이 변수 X_k가 서로 다른 factor F_1과 F_2에 대해 비슷한 loading 값을 가질 수도 있다. 즉, $\ell_{11} \fallingdotseq \ell_{12}$인 것이다. 이 말은 변수 X_k가 두 factor에 의해 비슷하게 영향을 받는다는 뜻이고, 변수들을 주로 영향받는 factor에 따라 분류한다고 했을 때 X_k를 F_1과 F_2 중 어디로 구분해야 할지 애매해진다.

Variable	F_1	F_2
X_1	ℓ_{11}	ℓ_{12}
$\vdots$	$\vdots$	$\vdots$
X_k	$\ell_{k1} = 0.36$	$\ell_{k2} = 0.40$
$\vdots$	$\vdots$	$\vdots$
X_p	ℓ_{p1}	ℓ_{p2}

Factor Rotation

Factor Rotation

$\hat{L} \in \mathbb{R}^{p \times m}$을 Principal Components/MLE 등을 통해 얻어진 estimated factor loading 행렬이라 하고, 임의의 orthogonal matrix $T \in \mathbb{R}^{m \times m}$를 정의하자. 그렇다면 orthogonal matrix의 성질에 따라 $TT' = T'T = I$이고, 따라서 $\hat{L}^* = \hat{L}T$를 정의하면

$$S_n = \hat{\Sigma} = \hat{L}^* \hat{L}^{*'} + \hat{\Psi} = (\hat{L}T)(T'\hat{L}') + \hat{\Psi} = \hat{L}\hat{L}' + \hat{\Psi}$$

이다. Orthogonal matrix의 곱에 residual matrix가 영향을 받지 않는 것이다. Orthogonal matrix는 벡터를 회전/반사하는 선형변환과 동치이다. 즉, factor loading 행렬의 회전/반사 변환은 결과에 어떤 영향도 미치지 않는 것이며, 이를 응용하면 더 나은 결과물을 얻을 수 있다.

Factor Rotation

Orthogonal Rotation

이해를 위해 아래 $m = 2$인 상황을 가정한 표를 2차원 좌표평면에 옮겨보자.

	Estimated factor loadings		Communalities
Variable	F_1	F_2	$\hat{h}_i^2$
X_1	.553	.429	.490
X_2	.568	.288	.406
X_3	.392	.450	.356
X_4	.740	−.273	.623
X_5	.724	−.211	.569
X_6	.595	−.132	.372

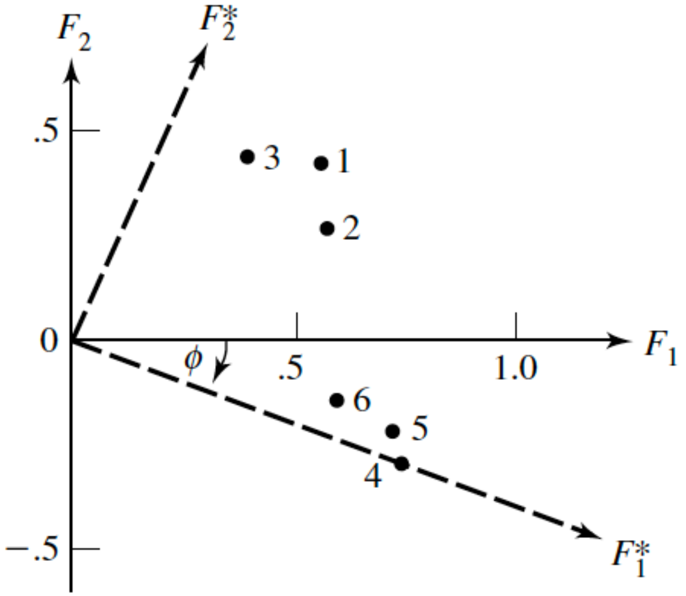

Figure 4-1 Orthogonal Rotation (1)

Factor Rotation

Orthogonal Rotation

이제 축을 회전시켜 회전된 축 위에 최소 하나의 점이 존재하도록 하자.

수정된 loading 값은 다음과 같다. 각 factor로의 구분이 훨씬 명확해졌다.

	Estimated factor loadings		Communalities	Estimated rotated factor loadings		Communalities
Variable	F_1	F_2	$\hat{h}_i^2$	F_1^*	F_2^*	$\hat{h}_i^{*2} = \hat{h}_i^2$
X_1	.553	.429	.490	.369	.594	.490
X_2	.568	.288	.406	.433	.467	.406
X_3	.392	.450	.356	.211	.558	.356
X_4	.740	−.273	.623	.789	.001	.623
X_5	.724	−.211	.569	.752	.054	.569
X_6	.595	−.132	.372	.604	.083	.372

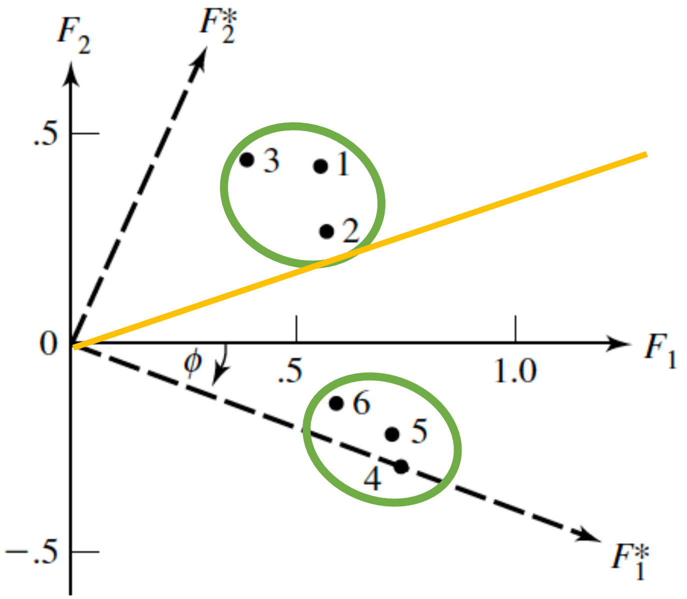

Figure 4-2 Orthogonal Rotation (2)

Factor Rotation

Orthogonal Rotation

이해를 위해 $m = 2$, 즉 common factor가 두 종류 뿐이라 하자.

이제 Orthogonal factor model의 행렬들은 다음과 같이 변환된다:

행렬 $\boldsymbol{F} \in \mathbb{R}^{2\times p}$일 것이며, $\hat{\boldsymbol{L}}^* = \hat{\boldsymbol{L}}\boldsymbol{T}$일 것이다. (단, $\hat{\boldsymbol{L}} \in \mathbb{R}^{p\times 2}, \boldsymbol{T} \in \mathbb{R}^{2\times 2}$)

이제 ϕ를 기존 common factor 벡터 F_i와 회전된 벡터 F_i^* 사이의 각도라고 하자. (counterclockwise) $m = 2$인 상황에서 기존 common factor들, 즉 F_1과 F_2를 서로 수직인 좌표축으로 생각할 수 있을 것이다. 이는 $\mathrm{Cov}(\boldsymbol{F}) = \boldsymbol{I}$이므로 $\mathrm{Cov}(F_1, F_2) = 0$인 것에 기인한다.

이제 행렬 $\boldsymbol{T}$는 다음과 같이 정의된다:

$$T = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$$

즉, $m = 2$인 상황은 **givens rotation**과 동일한 원리이다.

Factor Rotation

Orthogonal Rotation

$m = 3$인 경우를 생각해보자. 변수 X_i의 loading 값을 이용하여 $(\ell_{i1}, \ell_{i2}, \ell_{i3})$를 좌표공간에 나타내자. 이 중 가장 툭 튀어나온 좌표를 기준으로 $xy - plane$의 변환 각도 ϕ_1, $xz - plane$의 변환 각도 ϕ_2를 구하여 축 중 하나의 변환 정도를 계산하고, 이를 바탕으로 나머지 모든 축의 변환정도를 구할 수 있을 것이다. 이는 모든 축들이 서로 perpendicular한 상태가 변환 전후 모두 유지되어야 하기 때문이다.

가령 앞선 예시와 같이 좌표공간에 factor를 나타냈을 때, 기존 common factor가 $\boldsymbol{F} = span\{e_1, e_2, e_3\}$이고, 변환된 common factor가 $\boldsymbol{F}' = span\{e_1', e_2', e_3'\}$일 때, $\boldsymbol{F} \rightarrow \boldsymbol{F}'$의 기저변환을 수행하는 행렬을 구한다면 적절한 varimax 행렬 $\boldsymbol{T}$를 구할 수 있을 것이다. $m \geq 4$인 경우에도 마찬가지이다.

Factor Rotation

Varimax Rotation: Kaiser's Criterion

$\tilde{\ell}_{ij}^* = \hat{\ell}_{ij}^*/\hat{h}_i$이라 정의하자. 가능한 행렬 $\boldsymbol{T}$가 여러 개일 때, 어떤 것을 고르는 것이 좋을까? 이에 대한 Kaiser의 기준은 다음과 같다.

$$\boldsymbol{T} \coloneqq \max\left\{V = \frac{1}{p}\sum_{j=1}^{m}\left[\sum_{i=1}^{p}\tilde{\ell}_{ij}^{*4} - \left(\sum_{i=1}^{p}\tilde{\ell}_{ij}^{*2}\right)^2 /p\right]\right\}$$

이러한 V를 가능한한 크게 만드는 $\boldsymbol{T}$를 선택하면 작은 communality 값을 가진 변수들에 더 큰 가중치를 줄 수 있게 된다. $\hat{h}_i = \sum_{j=1}^{m}\hat{\ell}_{ij}^2$이므로, 즉 변수 X_i가 각 common factor F_i로부터 받는 영향력의 제곱합이다. 새로이 정의된 요인으로부터 받는 영향력이 작으면 변수를 구분짓기 어려우므로 위와 같은 기준을 세운 것이다.

V에 대한 해석

$\tilde{\ell}_{ij}^2 \coloneqq X$라 하자. 그렇다면 V가 분산을 의미, 즉

$$E(X^2) - [E(X)]^2$$

와 유사함을 알 수 있다. 결국 V를 극대화하겠다는 말은 X, 즉 loading 값의 분산을 극대화하겠다는 말과 동치이다.

Loading 값들의 분산이 커지게 되면 각 좌표들의 간격이 넓어지게 되고, 결국 점들을 구분하기 용이해질 것이다. 따라서 Kaiser's varimax criterion은 점들의 간격을 가능한한 넓히는 회전변환을 추천한다는 것이다.

Factor Rotation

Varimax Rotation

Varimax 회전 변환은 factor analysis가 Principal Components를 이용했을 때보다 Maximum Likelihood를 이용하였을 때 더 추천된다.

Maximum likelihood를 통해 loading 값들을 얻은 경우, $\hat{L}'\hat{\Psi}^{-1}\hat{L}$이 대각행렬이어야 한다는 uniqueness condition을 만족시키기 위해 초기 값들에 제약이 걸린다. 이러한 제약은 대각행렬이라는 조건으로 계산을 용이하게 만들어주기도 하지만, 해석하기 어려운 loading 값을 구해내게끔 만들기도 한다. 앞서 살펴봤든 우리는 Varimax Rotation을 통해 loading 값들의 구분을 좀 더 명확하게 할 수 있고, 따라서 위와 같은 상황에 추천되는 것이다.

Factor Rotation

Oblique Rotation

지금까지 살펴봤던 factor rotation은 common factor 행렬 F의 각 column F_i를 축으로 생각, 상호 orthogonal함이 유지되게끔 좌표축으로 표현하였다.

*실제로 $1 \le i \neq j \le m$일 때 $\mathrm{Cov}(F_i, F_j) = 0$이다.

그런데 oblique rotation은 원리는 동일하나, axis 간 orthogonality assumption의 깨짐을 허용한다. 즉 $\mathrm{Cov}(F_i^*, F_j^*) \neq 0$일 수도 있는 것이다.

이는 변수 간 상관관계 혹은 의존성을 인정하는 것으로서 좀 더 실전적인 상황에 유용한 방법이며, common factor의 수인 m을 가능한한 작게 만드는 것을 목적으로 한다.

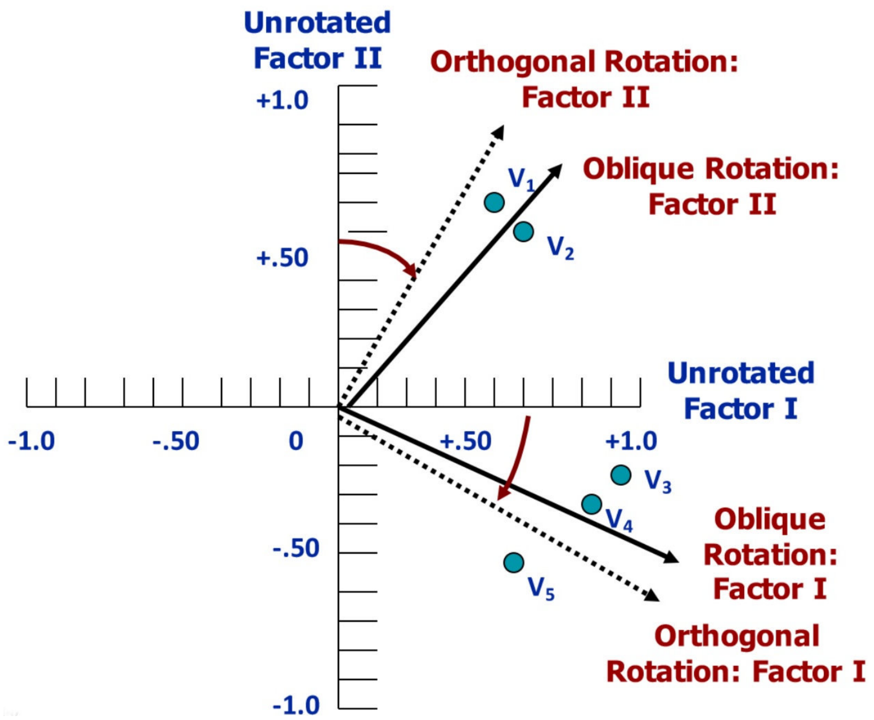

Figure 4-3 Orthogonal v. Oblique

5.Factor Scores

Factor Scores

Factor Scores

이제 Factor Rotation까지 완료하였다면 대부분의 factor analysis 자체는 끝이 났다. 그러나 정말 factor analysis가 적절하게 이루어졌다고 할 수 있을까? 다른 방법론들이 그러하듯, factor analysis도 결과물에 대한 평가 기준을 세우고, 이에 따라 점검하는 진단 과정이 필요하다. 이를 위해 Factor Score라는 새로운 개념을 도입하자.

$$\hat{\boldsymbol{f}}_j = \text{estimate of the values } \boldsymbol{f}_j \text{ attained by } \boldsymbol{F}_j \ (j\text{th case})$$

$\hat{\boldsymbol{f}}_j$는 우리가 알지 못하는 parameter를 대신할 추정값이 아니다. 그저 관측되지 않았을 뿐이다. Factor analysis의 각 factor $\boldsymbol{F}_j$는 실제로 존재하지만 데이터 상에서 겉으로 드러나지 않아 아직 눈치채지 못한 basis라고 생각하는 것이 좋다.

Factor Scores

Difficulty to find $\hat{\boldsymbol{f}}_j$

이제 $\hat{\boldsymbol{f}}_j$를 찾아보자. 그러나 이것은 매우 복잡하거나 때로는 불가능할 수도 있는데,

$$n(\boldsymbol{f}_j) + n(\boldsymbol{\epsilon}_j) \gg n(\mathbf{x}_j)$$

보통 factor analysis에서는 위와 같은 상황이 나타나기 때문이다. 그래서 우리는 이 상황을 타계할 해결책이 필요하고, 다음과 같다.

i. Treat the estimated values $\hat{\ell}_{ij}$, $\hat{\psi}_i$ as if they were the true values;

ii. Do not discriminate $\hat{\ell}^*_{ij}$(rotated) from $\hat{\ell}_{ij}$ (unrotated).

각 approach에 대한 해석

i. 만약 $\hat{\ell}_{ij}$, $\hat{\psi}_i$를 실제 값이 아니라 추정값으로 간주하게 되면, factor analysis 역시 어디까지 "추정"에 불과하게 된다. Factor analysis의 대전제가, latent factor를 찾아 이에 대해 span하는 것임을 생각하면 추정값으로 간주하는 것 자체가 common factor들이 latent하지 않다고 전제를 깨는 것과 다름이 없다.

ii. 이는 무조건적으로 rotated가 더 낫다고 생각하지 말자는 것이다. Factor score를 계산할 때 어떤 것을 이용하든 결과는 같기 때문이다.

Factor Scores

The Weighted Least Squared Method

$\boldsymbol{\mu}, \boldsymbol{L}, \boldsymbol{\Psi}$의 값이 알려진 다음 factor model을 생각해보자.

$$\mathbf{X}_{(p\times 1)} - \boldsymbol{\mu}_{(p\times 1)} = \mathbf{L}_{(p\times m)} \cdot \mathbf{F}_{(m\times 1)} + \boldsymbol{\epsilon}_{(p\times 1)}$$

$\epsilon' = [\epsilon_1, \epsilon_2, \cdots, \epsilon_p]$를 error라고 간주하자. $\mathrm{Cov}(\boldsymbol{\epsilon}) = \boldsymbol{\Psi}$의 대각성분 값인 $\mathrm{Var}(\epsilon_i) = \psi_i$에 대하여, $\psi_i \neq \psi_j$가 허용되므로, 이는 회귀분석의 OLS 모델로 간주할 수 있다. 그렇다면 $\hat{\boldsymbol{f}}_j$를 추정하기 위해 위 OLS 모델을 이용해보자. (Bartlett [2])

분산의 역수 가중치를 부여한 error의 제곱합은 다음과 같다:

$$\sum_{i=1}^{p} \frac{\epsilon_i^2}{\psi_i} = \boldsymbol{\epsilon}'\boldsymbol{\Psi}^{-1}\boldsymbol{\epsilon} = (\mathbf{x} - \boldsymbol{\mu} - \boldsymbol{L}\mathbf{f})'\boldsymbol{\Psi}^{-1}(\mathbf{x} - \boldsymbol{\mu} - \boldsymbol{L}\mathbf{f}).$$

Factor Scores

The Weighted Least Squared Method

당연히 error가 최소화될수록 좋은 값이므로, 이전 식을 최소화할 수 있는 $\mathbf{f}$를 최적값으로 잡을 수 있다:

$$\hat{\mathbf{f}} = \min_{\mathbf{f}} \sum_{i=1}^{p} \frac{\epsilon_i^2}{\psi_i}.$$

이를 계산하면 우리는 다음을 얻는다:

$$\hat{\mathbf{f}} = \left(\boldsymbol{L}'\boldsymbol{\Psi}^{-1}\boldsymbol{L}\right)^{-1}\boldsymbol{L}'\boldsymbol{\Psi}^{-1}(\mathbf{x} - \boldsymbol{\mu})$$

$$\hat{\mathbf{f}}_j = \left(\hat{\boldsymbol{L}}'\hat{\boldsymbol{\Psi}}^{-1}\hat{\boldsymbol{L}}\right)^{-1}\hat{\boldsymbol{L}}'\hat{\boldsymbol{\Psi}}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}})$$

$\hat{\boldsymbol{L}}$과 $\hat{\boldsymbol{\Psi}}$가 MLE를 통해 구해졌을 때, 이들은 $\hat{\boldsymbol{L}}'\hat{\boldsymbol{\Psi}}^{-1}\hat{\boldsymbol{L}} = \hat{\boldsymbol{\Delta}}$가 대각행렬이라는 uniqueness condition을 만족시켜야 한다.

지금까지의 내용을 요약하면 오른쪽과 같다.

Factor Scores Obtained by Weighted Least Squares from the Maximum Likelihood Estimates

$$\begin{aligned}\hat{\mathbf{f}}_j &= (\hat{\mathbf{L}}'\hat{\boldsymbol{\Psi}}^{-1}\hat{\mathbf{L}})^{-1}\hat{\mathbf{L}}'\hat{\boldsymbol{\Psi}}^{-1}(\mathbf{x}_j - \hat{\boldsymbol{\mu}}) \\ &= \hat{\boldsymbol{\Delta}}^{-1}\hat{\mathbf{L}}'\hat{\boldsymbol{\Psi}}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}}), \qquad j = 1, 2, \ldots, n\end{aligned}$$

Factor Scores

The Weighted Least Squared Method

만약 rotated된 상황에서는 어떻게 weighted least squared method를 통해 optimized factor score를 구할 수 있을까?

$\hat{\boldsymbol{L}}^* = \hat{\boldsymbol{L}}\boldsymbol{T}$로 rotation이 주어졌을 때, $\hat{\boldsymbol{L}}^* = \hat{\boldsymbol{L}}\boldsymbol{T}$로 치환하면 간단하게 $\hat{\mathbf{f}}_j^* = \boldsymbol{T}'\hat{\mathbf{f}}_j$임을 파악할 수 있다.

$\hat{\ell}_{ij}$가 principal component로 구해진 경우의 $\hat{\mathbf{f}}_j$

이 경우 각 error의 분산 역수 가중치를 주지 말고 factor score를 추정하면 된다. 간단히 말하면 $\boldsymbol{\Psi}^{-1}$를 식에서 제거하면 된다. 사실 이는 각 $\hat{\psi}_i$가 서로 동일하거나 거의 같다고 가정하는 것으로, principal component analysis는 각 변수의 기여도를 유사하게 설정하는 것에서 기인한다.

Factor Scores

The Regression Method

다시 $\boldsymbol{\mu}, \boldsymbol{L}, \boldsymbol{\Psi}$의 값이 알려진 다음 factor model을 생각해보자.

$$\mathbf{X}_{(p\times 1)} - \boldsymbol{\mu}_{(p\times 1)} = \mathbf{L}_{(p\times m)} \cdot \mathbf{F}_{(m\times 1)} + \boldsymbol{\epsilon}_{(p\times 1)}$$

위 모델에서 우리는 $X - \mu = LF + \epsilon \sim N_p(0, LL' + \Psi)$임을, 더 나아가 $(\mathbf{X} - \boldsymbol{\mu})$와 $\mathbf{F}$의 joint distribution이 $N_{m+p}(\mathbf{0}, \boldsymbol{\Sigma}^*)$를 따름 역시 알 수 있다.

여기서 $\boldsymbol{\Sigma}^*$는 다음과 같다:

$\mathrm{Cov}(X-\mu)$

$$\underset{(m+p)\times(m+p)}{\boldsymbol{\Sigma}^*} = \left[\begin{array}{c|c} \boldsymbol{\Sigma} = \underset{(p\times p)}{\mathbf{LL}'} + \boldsymbol{\Psi} & \underset{(p\times m)}{\mathbf{L}} \\ \hline \underset{(m\times p)}{\mathbf{L}'} & \underset{(m\times m)}{\mathbf{I}} \end{array}\right]$$

Factor Scores

The Regression Method

이를 바탕으로 $\boldsymbol{F}|\mathbf{x}$의 conditional multivariate normal distribution을 구하면 다음과 같은 parameter를 가진다:

$$\text{mean} = \mathrm{E}(\boldsymbol{F}|\mathbf{x}) = \boldsymbol{L}'\boldsymbol{\Sigma}^{-\mathbf{1}}(\mathbf{x}-\boldsymbol{\mu}) = \boldsymbol{L}'(\boldsymbol{L}\boldsymbol{L}'+\boldsymbol{\Psi})^{-\mathbf{1}}(\mathbf{x}-\boldsymbol{\mu})$$

$$\text{covariance} = \mathrm{Cov}(\boldsymbol{F}|\mathbf{x}) = \boldsymbol{I}-\boldsymbol{L}'\boldsymbol{\Sigma}^{-1}\boldsymbol{L} = \boldsymbol{I}-\boldsymbol{L}'(\boldsymbol{L}\boldsymbol{L}'+\boldsymbol{\Psi})^{-\mathbf{1}}\boldsymbol{L}$$

위의 mean 값은 다변량 회귀에서의 계수들과 같으므로, 이러한 계수들의 추정값으로부터 우리는 factor score를 추정할 수 있다:

$$\hat{\mathbf{f}}_j = \hat{\boldsymbol{L}}'\hat{\Sigma}^{-1}(\mathbf{x}_j-\bar{\mathbf{x}}) = \hat{\boldsymbol{L}}'\left(\hat{\boldsymbol{L}}\hat{\boldsymbol{L}}'+\hat{\Psi}\right)^{-1}(\mathbf{x}_j-\bar{\mathbf{x}}), \qquad j = 1,2,\cdots,n.$$

마찬가지로 $\hat{\boldsymbol{L}}^* = \hat{\boldsymbol{L}}\boldsymbol{T}$로 rotated된 상황의 경우라도, 이전과 동일하게 $\hat{\mathbf{f}}_j^* = \boldsymbol{T}'\hat{\mathbf{f}}_j$로 factor score를 구할 수 있다.

6.Strategy for Factor Analysis

Strategy for Factor Analysis

Strategy for Factor Analysis

결국 factor analysis에 있어 가장 중요한 것은 common factor의 개수를 몇 개로 잡을 것인가, 즉 $m \coloneqq n(\mathrm{f}_j)$를 몇 개로 설정할 것인가이다. Factor analysis와 PCA 모두 주요 목적은 변수의 개수를 줄이는 것이고, 따라서 m을 무작정 늘린다면 결국 분석의 의미가 사라지게 될 것이다. 결국 평가 기준을 세우고, 이에 따라 적정한 m을 잡는 것이 필요하다.

교재에서 제시하는 m을 설정하는 기준은 다음과 같다:

i. 샘플 분산의 설명 비율
ii. 데이터의 주제와 관련된 배경지식
iii. 결과의 합리성

PCA/MLE 선택, 회전 유무, 회전 알고리즘 등은 m을 설정하는 것에 비하면 사소한 요소일 뿐이다.

샘플 분산의 설명 비율

주어진 데이터 세트에서 선택한 common factor가 설명하는 분산의 비율을 의미한다. 점들의 분포를 common factor들만으로 얼마나 잘 설명할 수 있는지를 의미한다고 생각하면 된다. 그런데 설명 비율이 100%라고 무조건 좋은 것은 아닌데, 이는 $m \geq n(X_i)$이어서 차원 축소로서의 기능이 상실되거나, 데이터 셋에 overfitting 되었다는 의미이기 때문이다.

Strategy for Factor Analysis

Strategy for Factor Analysis

이후에는 다음과 같은 순서로 factor analysis를 진행하자.

i. Set an initial m value
ii. Principal component factor analysis (varimax rotation)
iii. Maximum likelihood factor analysis (varimax rotation)
iv. Compute factor scores of each results
v. Plot factor scores obtained for principal components against scores from maximum likelihood analysis
vi. Repeat ii ~ v for other m

*데이터셋의 크기가 큰 경우에는 데이터셋을 반으로 나누어 절반에 대해 factor analysis를 시행하고, 나머지 절반에 대해 검증을 진행하여 잘 이루어졌는지를 확인할 수 있다.

WOW Criterion

결국 factor analysis가 적절하게 시행되었다 하더라도, 각 common factor가 무엇을 의미하는지를 판단하는 것은 데이터에 대한 배경 지식과 실험자의 역량에 달렸다. 즉, 꽤나 주관적인 것이다. 따라서 분석 결과가 매우 적절하다고 판단되는 경우, 즉 "WOW"라는 감탄사가 나오는 경우라면 그 분석은 정량적인 사후검정을 거치지 않아도 잘 이루어진 분석이라고 할 수 있을 것이다.

END

References

Figure

[1] Johnson, R. A., & Wichern, D. W. (2015). *Applied multivariate statistical analysis* (6th ed.)(p.483). PEARSON.

[2] Johnson, R. A., & Wichern, D. W. (2015). *Applied multivariate statistical analysis* (6th ed.)(p.483). PEARSON.

[3] Johnson, R. A., & Wichern, D. W. (2015). *Applied multivariate statistical analysis* (6th ed.)(p.490). PEARSON.

[4] Johnson, R. A., & Wichern, D. W. (2015). *Applied multivariate statistical analysis* (6th ed.)(p.492). PEARSON.