多元统计分析

第8讲 因子分析(I)

Johnson & Wichern Ch9.1-9.3

统计学研究中心 邓婉璐
wanludeng@tsinghua.edu.cn

2018-2019春季学期

Motivating Example

- Here are some scores for students from high school.

满分100	数学	物理	化学	历史	政治	语文	总分
A	98	97	93	73	70	80	511
B	75	71	80	95	94	96	511

- How would you like to describe the student A and B?
- Correlation matrix:

R	数学	物理	化学	历史	政治	语文
数学	1	0. 44	0. 41	0. 28	0. 29	0. 25
物理		1	0. 35	0. 16	0. 19	0. 18
化学			1	0. 31	0. 32	0. 32
历史				1	0. 61	0. 47
政治					1	0. 46
语文						1

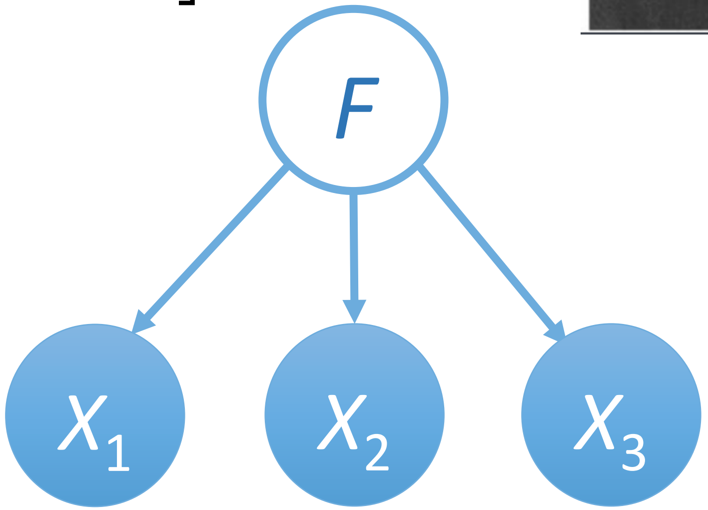
Motivating Example

- Charles Spearman(1863-1945): 1904, Consider children's exam performance in X_1 = Classics, X_2 = French, X_3 = English, with observation correlation matrix

$$R = \begin{bmatrix} 1 & 0.83 & 0.78 \\ 0.83 & 1 & 0.67 \\ 0.78 & 0.67 & 1 \end{bmatrix}$$

- Model:

$$X_1 = l_1 F + \varepsilon_1$$
$$X_2 = l_2 F + \varepsilon_2$$
$$X_3 = l_3 F + \varepsilon_3$$



Outline

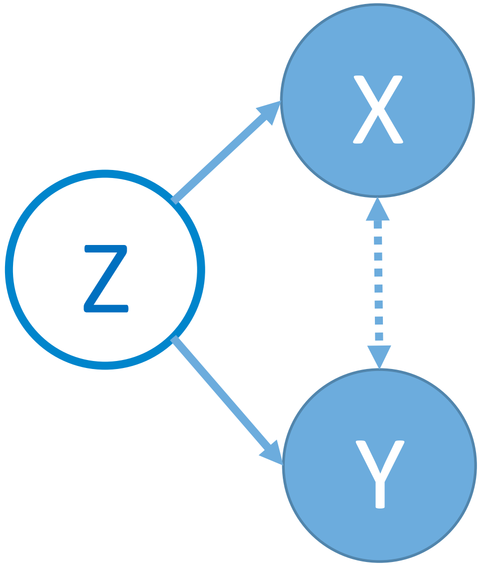
- Introduction and Model
- Methods for Estimation
  - PC method
  - MLE method
- Explanation – Rotation
- Factor Scores
  - Weighted LSE Method
  - Regression Method

Introduction and Model

Overview of Factor Analysis

- Early development in psychometrics by Karl Pearson, Charles Spearman, etc
- To describe the covariance structure among many variables with a few unobservable or latent variables called factors
  - Reduction: reduce high dimension data to a few variables
  - Interpretation: explain the covariance of observed variables with latent factors

$$\underset{(p\times p)}{\boldsymbol{\Sigma}} = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{bmatrix}$$



Orthogonal Factor Model (I)

- The observable random vector **X**, with *p* components, with mean and covariance

- X is linearly dependent upon a few common factors and specific factors, with

loading 也叫因子载荷

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

or

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

这里的因子类似设计变量，也可以是随机的

Orthogonal Factor Model (II)

- Assumptions continued

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

一般模型

$$E(\mathbf{F}) = 0, \quad \mathrm{Cov}(\mathbf{F}) = E(\mathbf{F}\mathbf{F}') = \underset{(m\times m)}{\mathbf{I}}$$

$$E(\boldsymbol{\varepsilon}) = 0, \quad \mathrm{Cov}(\boldsymbol{\varepsilon}) = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \underset{(p\times p)}{\boldsymbol{\Psi}}$$

$$\boldsymbol{\Psi} = \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_p \end{bmatrix}$$

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

每个因子正交，

误差项正交

Covariance Structure Implied

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

$$= (\mathbf{LF})(\mathbf{F}'\mathbf{L}') + \boldsymbol{\varepsilon}(\mathbf{F}'\mathbf{L}') + (\mathbf{LF})\boldsymbol{\varepsilon}' + \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'$$

$$\boldsymbol{\Sigma} = \mathrm{Cov}(\mathbf{X}) = E(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})'$$

$$= \mathbf{L}E(\mathbf{FF}')\mathbf{L}' + E(\boldsymbol{\varepsilon}\mathbf{F}')\mathbf{L}' + \mathbf{L}E(\mathbf{F}\boldsymbol{\varepsilon}') + E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')$$

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

$$\sigma_{ii} = \boxed{l_{i1}^2 + l_{i2}^2 + \cdots + l_{im}^2} + \boxed{\psi_i}$$

$$= h_i^2 + \psi_i$$

Communality + Specific variance

$$\sigma_{ik} = \mathrm{Cov}(X_i, X_k) = \mathbf{l_i}'\mathbf{l}_k$$

$$= l_{i1}l_{k1} + \cdots + l_{im}l_{km}$$

$$(\mathbf{X}-\boldsymbol{\mu})\mathbf{F}' = (\mathbf{LF}+\boldsymbol{\varepsilon})\mathbf{F}'$$

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

$$\mathrm{Cov}(X_i, F_j) = l_{ij}$$

Covariance

Example (Textbook Ex9.1)

分析的时候却是只需要考虑方差就好了

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

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

Example (Textbook Ex9.1)

$$\mathbf{L} = \begin{bmatrix} \ell_{11} & \ell_{12} \\ \ell_{21} & \ell_{22} \\ \ell_{31} & \ell_{32} \\ \ell_{41} & \ell_{42} \end{bmatrix} = \begin{bmatrix} 4 & 1 \\ 7 & 2 \\ -1 & 6 \\ 1 & 8 \end{bmatrix},$$

$$\boldsymbol{\Psi} = \begin{bmatrix} \psi_1 & 0 & 0 & 0 \\ 0 & \psi_2 & 0 & 0 \\ 0 & 0 & \psi_3 & 0 \\ 0 & 0 & 0 & \psi_4 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

$$h_1^2 = \ell_{11}^2 + \ell_{12}^2 = 4^2 + 1^2 = 17$$

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

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

Understanding the model - Reduction

- How many parameters are there in a covariance matrix?
- How many parameters are there in the orthogonal factor model?
- What is the maximum number of common factors?

Limitation of FA model: Not all covariance matrix can be factored as $\mathbf{LL}'+\mathbf{\Psi}$, where the number of factors $m << p$

See Example 9.2 in textbook (maybe the solution exists mathematically, but not statistically; e.g. correlation > 1 or variance < 0, – (ultra) Heywood case (Heywood 1931))

Understanding the model - Nonidentifiability

- Consider orthogonal matrix **T**

$$\mathbf{X}-\boldsymbol{\mu}=\mathbf{L}\mathbf{F}+\boldsymbol{\varepsilon}$$

Check model assumptions

$$=\mathbf{L}(\mathbf{T}\mathbf{T}')\mathbf{F}+\boldsymbol{\varepsilon}$$

$$=(\mathbf{L}\mathbf{T})(\mathbf{T}'\mathbf{F})+\boldsymbol{\varepsilon}$$

增加一个正交矩阵之后还是因子分解

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

$$\mathbf{X}-\boldsymbol{\mu}=\mathbf{L}^*\mathbf{F}^*+\boldsymbol{\varepsilon}$$

L is not unique!

- Since the model doesn't change, we'll later use this in two ways
  - ✧ To "rotate" the factors to make them more interpretable
  - ✧ To assist in optimization for maximum likelihood estimation.

Understanding the model - Scale Invariant

- Consider diagonal matrix **C**

$$\mathbf{X} = \mu + \mathbf{LF} + \varepsilon$$

$$Y = CX = C\mu + C\mathbf{LF} + C\varepsilon$$
$$= \mu_C + L_C F + \varepsilon_C$$

The structures are not affected by choices of **C**

$Var(\varepsilon_C) = C\Psi C'$ is diagonal since C is diagonal.

- So, in this sense, FA is unaffected by rescaling of the variables.

Targets of Model

- Suppose $\mathbf{x_1}, \mathbf{x_2}, \cdots, \mathbf{x_n}$ represent *n* independent drawings from some p-dimensional population, with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.
- Sample covariance matrix **S,** sample correlation matrix **R**
- Objective: find $\hat{\mathbf{L}}$ and $\hat{\boldsymbol{\Psi}}$, with $\mathbf{S} \approx \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}}$

What is the meaning of the latent factors?

✧ L

✧ Rotation

What are the latent factors?

✧ F

Principle Component Approach for Estimating Loadings

Principal Component Approach

- Objective: find $\hat{\mathbf{L}}$ and $\hat{\mathbf{\Psi}}$, with $\mathbf{S} \approx \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\mathbf{\Psi}}$
- Intuitively, we may use PCA / spectral decomposition:

$$\Sigma = [\sqrt{\lambda_1}e_1 \vdots \sqrt{\lambda_2}e_2 \vdots \cdots \vdots \sqrt{\lambda_p}e_p] \begin{bmatrix} \sqrt{\lambda_1}e_1' \\ \sqrt{\lambda_2}e_2' \\ \vdots \\ \sqrt{\lambda_p}e_p' \end{bmatrix} = \underset{(p\times p)}{L_0}\underset{(p\times p)}{L_0'} + \underset{(p\times p)}{0} = \underset{(p\times p)}{L_0}\underset{(p\times p)}{L_0'}$$

The spectral decomposition is not useful!
common factors = # variables

Principal Component Approach

- When the last *p-m* eigenvalues are small, neglect the contribution of the corresponding eigenvalue-eigenvector pairs

$$\Sigma \approx [\sqrt{\lambda_1} e_1 \vdots \sqrt{\lambda_2} e_2 \vdots \cdots \vdots \sqrt{\lambda_m} e_m] \begin{bmatrix} \sqrt{\lambda_1} e_1' \\ \sqrt{\lambda_2} e_2' \\ \vdots \\ \sqrt{\lambda_m} e_m' \end{bmatrix} = \underset{(p\times m)}{L} \underset{(m\times p)}{L'}$$

What is $\tilde{\boldsymbol{\Psi}}$?

What is communality $\tilde{h}_i^2$?

Given sample covariance matrix S or sample correlation matrix R

$$\tilde{L} = [\sqrt{\hat{\lambda}_1} \hat{e}_1 \vdots \sqrt{\hat{\lambda}_2} \hat{e}_2 \vdots \cdots \vdots \boxed{\sqrt{\hat{\lambda}_m} \hat{e}_m}]$$

Factor loadings

Example (Textbook Ex9.1)

$$\Sigma = L_0 L_0' \approx LL' + \boxed{\Psi}$$

For illustration, the numbers are rounded.

$$\begin{bmatrix} 19 & \boxed{30} & 2 & 12 \\ 30 & 57 & 5 & 23 \\ 2 & 5 & 38 & 47 \\ 12 & 23 & 47 & 68 \end{bmatrix} = \begin{bmatrix} -2.5 & 3.3 & 1.3 & -0.4 \\ -4.8 & 5.8 & -0.8 & 0.0 \\ -5.1 & -3.3 & -0.3 & -1.0 \\ -7.8 & -2.4 & 0.2 & 0.8 \end{bmatrix} \times \begin{bmatrix} -2.5 & -4.8 & -5.1 & -7.8 \\ 3.3 & 5.8 & -3.3 & -2.4 \\ 1.3 & -0.8 & -0.3 & 0.2 \\ -0.4 & 0.0 & -1.0 & 0.8 \end{bmatrix}$$

$$= \begin{bmatrix} -2.5 & 3.3 \\ -4.8 & 5.8 \\ -5.1 & -3.3 \\ -7.8 & -2.4 \end{bmatrix} \times \begin{bmatrix} -2.5 & -4.8 & -5.1 & -7.8 \\ 3.3 & 5.8 & -3.3 & -2.4 \end{bmatrix} + \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

$$\approx \begin{bmatrix} 17 & \boxed{31} & 2 & 12 \\ 31 & 56 & 5 & 23 \\ 2 & 5 & 37 & 48 \\ 12 & 23 & 48 & 67 \end{bmatrix} + \boxed{\begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}$$

Principal Component Approach

- Given sample covariance matrix S or sample correlation matrix R

$$\tilde{L} = [\sqrt{\hat{\lambda}_1}\hat{e}_1 \vdots \sqrt{\hat{\lambda}_2}\hat{e}_2 \vdots \cdots \vdots \boxed{\sqrt{\hat{\lambda}_m}\hat{e}_m}]$$

Factor loadings

- The specific variances may be taken to be the diagonal elements of $\boxed{S - \tilde{\mathbf{L}}\tilde{\mathbf{L}}'}$ Not diagonal

- Note: the estimated loadings for a given factor do not change as the number of factors is increased.

Example (Textbook Ex9.1)

$$\Sigma = L_0 L_0' \approx LL' + \boxed{\Psi}$$

For illustration, the numbers are rounded.

$$\begin{matrix} 19 & 30 & 2 & 12 \\ 30 & 57 & 5 & 23 \\ 2 & 5 & 38 & 47 \\ 12 & 23 & 47 & 68 \end{matrix} \approx \begin{matrix} -2.5 \\ -4.8 \\ -5.1 \\ -7.8 \end{matrix} \times \begin{matrix} -2.5 & -4.8 & -5.1 & -7.8 \end{matrix} + \begin{matrix} 13 & 0 & 0 & 0 \\ 0 & 34 & 0 & 0 \\ 0 & 0 & 12 & 0 \\ 0 & 0 & 0 & 7 \end{matrix}$$

$$\approx \begin{matrix} -2.5 & 3.3 \\ -4.8 & 5.8 \\ -5.1 & -3.3 \\ -7.8 & -2.4 \end{matrix} \times \begin{matrix} -2.5 & -4.8 & -5.1 & -7.8 \\ 3.3 & 5.8 & -3.3 & -2.4 \end{matrix} + \begin{matrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{matrix}$$

$$\approx \begin{matrix} -2.5 & 3.3 & 1.3 \\ -4.8 & 5.8 & -0.8 \\ -5.1 & -3.3 & -0.3 \\ -7.8 & -2.4 & 0.2 \end{matrix} \times \begin{matrix} -2.5 & -4.8 & -5.1 & -7.8 \\ 3.3 & 5.8 & -3.3 & -2.4 \\ 1.3 & -0.8 & -0.3 & 0.2 \end{matrix} + \begin{matrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{matrix}$$

Selection of *m*

Matrix Approximation

- Similar to PCA
- Consider the residual matrix $\mathbf{S} - (\tilde{\mathbf{L}}\tilde{\mathbf{L}}' + \tilde{\boldsymbol{\Psi}})$
- It can be shown that

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

- Consequently, a small value for the sum of the squares of the neglected eigenvalues implies a small value for the sum of the squared errors of approximation.

Selection of *m*: another perspective

- What is the total variance in **X**?

$$tr(S) = s_{11} + \cdots + s_{pp}$$

- What is the contribution of common factor *i* to the total variance?

$$l_{1i}^2 + \cdots + l_{pi}^2 = \hat{\lambda}_i$$

- Proportion of variance explained by the common factors

$$\frac{\hat{\lambda}_1 + \cdots + \hat{\lambda}_m}{tr(S)}$$

Example: Consumer Preference

$$\begin{array}{l} \text{Taste} \\ \text{Good buy for money} \\ \text{Flavor} \\ \text{Suitable for snack} \\ \text{Provides lots of energy} \end{array} \begin{bmatrix} 1 & 0.02 & \boxed{0.96} & 0.42 & 0.01 \\ & 1 & 0.13 & 0.71 & \boxed{0.85} \\ & & 1 & 0.50 & 0.11 \\ & & & 1 & \boxed{0.79} \\ & & & & 1 \end{bmatrix}$$

In fact, from the view of PCA, only 2 eigenvalues of sample correlation matrix ***R*** exceeds 1.

Finally, we choose m = 2.

What is the *m*?

Variable	Estimated factor loadings F_1	Estimated factor loadings F_2	Communalities h_i^2	Specific variances ψ_i
Taste	0.56	0.82	0.98	0.02
Good buy for money	0.78	−0.53	0.88	0.12
Flavor	0.65	0.75	0.98	0.02
Suitable for snack	0.94	−0.10	0.89	0.11
Provides lots of energy	0.80	−0.54	0.93	0.07
Eigenvalues	2.85	1.81		
Cumulative proportion of total (standardized) sample variance	0.571	0.932		

Table 9.1

Principle Factor Approach for Estimating Loadings

Principal Factor Approach

- Intuitive idea: the common factors should account for the off-diagonal elements, as well as the communality portions of the diagonal elements

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

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

Iteration for optimization:

Initial $\tilde{\boldsymbol{\Psi}}$

Find $\tilde{\mathbf{L}}$, the largest m eigenvectors of the eigen decomposition of $\mathbf{S} - \tilde{\boldsymbol{\Psi}}$

$$\tilde{\boldsymbol{\Psi}} = diag(\mathbf{S} - \tilde{\mathbf{L}}\tilde{\mathbf{L}}')$$

Discussions

- Choice of initial estimates of specific variances
- Some of the eigenvalues of $\mathbf{S} - \tilde{\mathbf{\Psi}}$ may be negative
- Communality may exceed total variance, Heywood case
- Reasonable suggestion of how to initial $\tilde{\mathbf{\Psi}}$

$$h_i^{*2} = 1 - \psi_i^* = 1 - \frac{1}{r^{ii}}$$

r^{ii} is the i-th diagonal element of $\mathbf{R}^{-1}$

Maximum Likelihood Approach for Estimating Loadings

Maximum Likelihood Approach

- Assumption: the common factors and the specific factors are jointly normally distributed

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

$$\mathbf{F} \sim N_m(0, \mathbf{I})$$

$$\boldsymbol{\varepsilon} \sim N_p(0, \boldsymbol{\Psi})$$

$$F \perp \varepsilon$$

$$\Rightarrow$$

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

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

subject to $\mathbf{L}'\boldsymbol{\Psi}^{-1}\mathbf{L} = \boldsymbol{\Delta}$

Maximum Likelihood Approach

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

$$= (2\pi)^{-\frac{(n-1)p}{2}} |\boldsymbol{\Sigma}|^{-\frac{(n-1)}{2}} \exp\{-\frac{1}{2} tr[\boldsymbol{\Sigma}^{-1}(\sum_{j=1}^{n}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})')]\}$$

$$\times (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\{-\frac{n}{2}(\bar{\mathbf{x}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu})\}$$

The model depends on $\mathbf{L}$ and $\boldsymbol{\Psi}$ through $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}$
It is not well defined because of multiplicity of choices of $\mathbf{L}$

Impose computationally convenient uniqueness condition:

$$\mathbf{L}'\boldsymbol{\Psi}^{-1}\mathbf{L} = \boldsymbol{\Delta}, \quad \boldsymbol{\Delta} \text{ is a diagonal matrix}$$

Maximum Likelihood Approach

Res 9.1

Let $X_1,\ldots,X_n$ be a random sample from $N_p(\mu,\Sigma)$, where $\Sigma = LL' + \Psi$ is the covariance matrix for the m common factor model. The maximum likelihood estimators $\hat{L}$, $\hat{\Psi}$, and $\hat{\mu}$ subject to $\hat{L}\hat{\Psi}^{-1}\hat{L}'$ being diagonal.

Then, the MLE of the communalities are

共有方差

$$\hat{h}_i^2 = \hat{l}_{i1}^2 + \hat{l}_{i2}^2 + \cdots + \hat{l}_{im}^2, \text{ for } i = 1,2,\cdots,p$$

so

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

Standardization

- If the variables are standardized so that $Z=V^{-1/2}(X-\mu)$
- What is the covariance matrix ρ?

$$\boldsymbol{\rho} = \mathbf{V}^{-1/2}\boldsymbol{\Sigma}\mathbf{V}^{-1/2} = (\mathbf{V}^{-1/2}\mathbf{L})(\mathbf{V}^{-1/2}\mathbf{L})' + \mathbf{V}^{-1/2}\boldsymbol{\Psi}\mathbf{V}^{-1/2}$$

- Thus, we have a factorization of ρ:

$$\mathbf{L}_z = \mathbf{V}^{-1/2}\mathbf{L}, \ \boldsymbol{\Psi}_z = \mathbf{V}^{-1/2}\boldsymbol{\Psi}\mathbf{V}^{-1/2}$$

- The MLE of ρ is

$$\hat{\boldsymbol{\rho}} = (\hat{\mathbf{V}}^{-1/2}\hat{\mathbf{L}})(\hat{\mathbf{V}}^{-1/2}\hat{\mathbf{L}})' + \hat{\mathbf{V}}^{-1/2}\hat{\boldsymbol{\Psi}}\hat{\mathbf{V}}^{-1/2}$$

$$= \hat{\mathbf{L}}_z\hat{\mathbf{L}}_z' + \hat{\boldsymbol{\Psi}}_z$$

Question:

Why?

What is $\hat{\mathbf{V}}^{-1/2}$?

Does PC approach have similar property?

Note:

- The MLE method could produce very different results when m→ m+1
- The MLE method can also experience difficulties with Heywood cases

Example: Stock-price data

Variable	Maximum likelihood			Principal components		
	Estimated factor loadings		Specific variances	Estimated factor loadings		Specific variances
	F_1	F_2	ψ_i	F_1	F_2	ψ_i
J P Morgan	0.115	0.755	0.42	0.732	−0.437	0.27
Citibank	0.322	0.788	0.27	0.831	−0.280	0.23
Wells Fargo	0.182	0.652	0.54	0.726	−0.374	0.33
Royal Dutch Shell	1.000	−0.000	0.00	0.605	0.694	0.15
ExxonMobil	0.683	−0.032	0.53	0.563	0.719	0.17
Cumulative proportion of total (standardized) sample variance	0.323	0.647		0.487	0.769	

Table 9.3

Discussion:

- Are the columns orthogonal?
- Estimated value (specific variances)
- % of total variance explained?

Homework:
princomp() / factpc() (library(mvstats))
factanal()

A Large Sample Test for the Number of Common Factors

- Normality Assumption: the common factors and the specific factors are jointly normally distributed

$$H_0: \underset{(p\times p)}{\Sigma} = \underset{(p\times m)}{\mathbf{L}}\underset{(m\times p)}{\mathbf{L}'} + \underset{(p\times p)}{\Psi}, \text{ subject to } \mathbf{L}'\Psi^{-1}\mathbf{L} = \Delta$$

$$H_1: \quad \Sigma \text{ any other positive definte matrix}$$

- Likelihood Ratio test $-2\ln\Lambda = -2\ln\left[\dfrac{\text{maximum likelihood under } H_0}{\text{maximized likelihood}}\right]$

A Large Sample Test for the Number of Common Factors

- Under null:

$$\hat{\mu} = \bar{\mathbf{x}}, \qquad \hat{\boldsymbol{\Sigma}} = \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}}$$

the maximized likelihood $\propto$

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

- Under alternative:

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}, \ \hat{\boldsymbol{\Sigma}} = \frac{n-1}{n}\mathbf{S}, \text{or } \mathbf{S}_\text{n}$$

$$\text{the maximized likelihood} \propto |\mathbf{S}_n|^{-n/2} e^{-np/2}$$

$$-2\ln\Lambda = -2\ln\left[\frac{\text{maximum likelihood under } \text{H}_0}{\text{maximized likelihood}}\right] = n\ln\left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\mathbf{S}_n|}\right)$$

Degree of Freedom :

$$\nu \text{-} \nu_0 = \frac{1}{2}p(p+1) - p(m+1) + \frac{1}{2}m(m-1)$$

Key Points

Orthogonal Factor Model (motivation, assumptions, concepts)

- Latent factor variables, dimension reduction
- Zero mean, uncorrelated
- Common factor, specific factor, communality
- Not unique / non-identifiable, scale invariant

Methods of Estimation

- PC method
  - Factor loading
  - How to determine m (the number of common factors)
- Principal factor solution (not required)
- Maximum likelihood method
  - Assumption: normality
  - Restriction for uniqueness
  - How to determine m (the number of common factors)
  - Standardized version
  - Comparison to PC method