

# Factor analysis

ZIH-SYUAN, JIAN

National Central University

September 23, 2018

# Outline of section

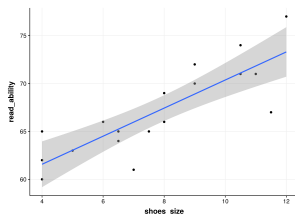
- 1 Introduction
  - Motivation
  - Purpose of factor analysis
- 2 Modeling
  - The Orthogonal Factor Model
- 3 Method
  - Principal Component
  - MLE
- 4 Rotation
- 5 Factor Score
- 6 Factor Analysis in R(Python)
- 7 Extended reading

# Motivation example (Using a Regression example)

For children in elementrays school

**Observed variable:** Shoe size and reading ability

**latent variable:** age



# Purpose of factor analysis

- Dimension reduction (it is not the most important in factor analysis)
- Describe the covariance relationship among many observable variables in terms of a few underlying, but unobservable (latent) variable.

# Outline of section

- 1 Introduction
  - Motivation
  - Purpose of factor analysis
- 2 Modeling
  - The Orthogonal Factor Model
- 3 Method
  - Principal Component
  - MLE
- 4 Rotation
- 5 Factor Score
- 6 Factor Analysis in R(Python)
- 7 Extended reading

# The Orthogonal Factor Model

The observable random vector  $\mathbf{X}$ , with components has mean  $\mu$  and covariance matrix  $\Sigma$ .  $\mathbf{X}$  is linearly dependent upon a few unobservable random variables  $F_1, F_2, \dots, F_m$  called **common factor**

$$X_1 - \mu_1 = \ell_{11}F_1 + \ell_{12}F_2 + \dots + \ell_{1m}F_m + \varepsilon_1$$

$$X_2 - \mu_2 = \ell_{21}F_1 + \ell_{22}F_2 + \dots + \ell_{2m}F_m + \varepsilon_2$$

$$\vdots$$

$$X_p - \mu_p = \ell_{p1}F_1 + \ell_{p2}F_2 + \dots + \ell_{pm}F_m + \varepsilon_p$$

in matrix notation

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

Assume that

$$E(\mathbf{F}) = \mathbf{0}_{(m \times 1)}, \text{Cov}(\mathbf{F}) = E(\mathbf{FF}') = \mathbf{I}_{(m \times m)}$$

$$E(\boldsymbol{\varepsilon}) = \mathbf{0}_{(p \times 1)}, \text{Cov}(\boldsymbol{\varepsilon}) = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \boldsymbol{\Psi}_{(p \times p)}$$

where  $\boldsymbol{\Psi}$  is a diagonal matrix.

So the covariance structure for  $\mathbf{X}$  From this model

$$\begin{aligned}
 \Sigma &= \text{Cov}(\mathbf{X}) = E((\mathbf{X} - \mu)(\mathbf{X} - \mu)') \\
 &= \mathbf{L}E(\mathbf{F}\mathbf{F}')\mathbf{L}' + \cancel{E(\boldsymbol{\varepsilon}\mathbf{F}')\mathbf{L}'}^0 + \cancel{\mathbf{L}E(\mathbf{F}\boldsymbol{\varepsilon}')}'^0 + E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') \\
 &= \mathbf{L}\mathbf{L}' + \Psi
 \end{aligned}$$

$$\text{Cov}(\boldsymbol{\varepsilon}, \mathbf{F}) = E(\boldsymbol{\varepsilon}\mathbf{F}') = 0,$$

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

- Communality  $h_i^2$ : portion of  $\text{Var}(X_i)$  contributed by the  $m$  common factors.
- Specific variance  $\psi_i$ : portion of  $\text{Var}(X_i)$  due to specific factor.

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



# Outline of section

- 1 Introduction
  - Motivation
  - Purpose of factor analysis
- 2 Modeling
  - The Orthogonal Factor Model
- 3 Method
  - Principal Component
  - MLE
- 4 Rotation
- 5 Factor Score
- 6 Factor Analysis in R(Python)
- 7 Extended reading

# Principal Component

The spectral decomposition of provides is with one factoring of the covariance matrix  $\Sigma$ . Let  $\Sigma$  have eigenvalue-eigenvector pairs  $(\lambda_i, e_i)$  with  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ . Then

$$\begin{aligned}\Sigma &= \lambda_1 e_1 e_1' + \lambda_2 e_2 e_2' + \dots + \lambda_p e_p e_p' \\ &= \begin{bmatrix} \sqrt{\lambda_1} e_1 & \sqrt{\lambda_2} e_2 & \dots & \sqrt{\lambda_p} e_p \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} e_1' \\ \sqrt{\lambda_2} e_2' \\ \vdots \\ \sqrt{\lambda_p} e_p' \end{bmatrix}\end{aligned}$$

The above form is not useful since  $m = p$ .

Most covariance matrices cannot be factored as  $\mathbf{LL}'$  when  $m$  is much less than  $p$ ;

So allowing for specific factors, we find that the approximation becomes

$$\begin{aligned}\Sigma &\doteq \mathbf{LL}' + \psi \\ &= \begin{bmatrix} \sqrt{\lambda_1}e_1 & \sqrt{\lambda_2}e_2 & \cdots & \sqrt{\lambda_m}e_m \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1}e'_1 \\ \sqrt{\lambda_2}e'_2 \\ \vdots \\ \sqrt{\lambda_p}e'_m \end{bmatrix} \\ &\quad + \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_p \end{bmatrix}\end{aligned}$$

Use **S** or **R** in place of  $\Sigma$

# How to choose $m$ ?

The  $m$  is chosen such that a “suitable proportion” of the total sample variance is explained

$$\left( \begin{array}{c} \text{Proportion of total} \\ \text{sample variance} \\ \text{due to } j\text{th factor} \end{array} \right) = \left\{ \begin{array}{ll} \frac{\hat{\lambda}_j}{s_{11} + s_{22} + \cdots + s_{pp}} & \text{for a factor analysis of } \mathbf{S} \\ \frac{\hat{\lambda}_j}{p} & \text{for a factor analysis of } \mathbf{R} \end{array} \right. \quad (9-20)$$

# Estimation by MLE

The common factors  $\mathbf{F}$  and specific factors  $\varepsilon$  are assumed to be normally distributed:  $\mathbf{F} \sim \mathcal{N}_m(\mathbf{0}, \mathbf{I}_m)$  and  $\varepsilon \sim \mathcal{N}_p(\mathbf{0}, \Psi)$

# Outline of section

- 1 Introduction
  - Motivation
  - Purpose of factor analysis
- 2 Modeling
  - The Orthogonal Factor Model
- 3 Method
  - Principal Component
  - MLE
- 4 Rotation
- 5 Factor Score
- 6 Factor Analysis in R(Python)
- 7 Extended reading

# Rotation

- Factor rotation = Orthogonal transformation of the factors
- Different orthogonal transformation lead to the same estimated covariance matrix, residual matrix, specific variances, and communalities

$$\mathbf{L}^* = \mathbf{L}\mathbf{T} \text{ then } \mathbf{L}\mathbf{L}' + \Psi = \mathbf{L}\mathbf{T}\mathbf{T}'\mathbf{L}' + \Psi = \mathbf{L}^*\mathbf{L}^{*'} + \Psi$$

- Different orthogonal transformation lead to different meanings/interpretations of factors
- Select the one with most meaningful interpretation

- Varimax rotation

Kaiser [9] has suggested an analytical measure of simple structure known as the *varimax* (or normal varimax) *criterion*. Define  $\tilde{\ell}_{ij}^* = \tilde{\ell}_{ij} / h_i$  to be the rotated coefficients scaled by the square root of the communalities. Then the (normal) varimax procedure selects the orthogonal transformation  $\mathbf{T}$  that makes

$$V = \frac{1}{p} \sum_{j=1}^m \left[ \sum_{i=1}^p \tilde{\ell}_{ij}^{*4} - \left( \sum_{i=1}^p \tilde{\ell}_{ij}^{*2} \right)^2 / p \right] \quad (9-45)$$

as large as possible.

$$V \propto \sum_{j=1}^m \left( \text{variance of squares of (scaled) loadings for } j\text{th factor} \right) \quad (9-46)$$

$$V = \sum_{j=1}^m \left\{ \frac{1}{p} \sum_{i=1}^p (\tilde{\ell}_{ij}^*)^4 - \left( \frac{1}{p} \sum_{i=1}^p \tilde{\ell}_{ij}^{*2} \right)^2 \right\}$$

\* Idea =  $L = \begin{bmatrix} \begin{bmatrix} \times \\ \times \\ \times \\ \times \\ \times \end{bmatrix} & \begin{bmatrix} \times \\ \times \\ \times \\ \times \\ \times \end{bmatrix} & \dots & \begin{bmatrix} \times \\ \times \\ \times \\ \times \\ \times \end{bmatrix} \end{bmatrix}_{p \times m} \longrightarrow L^* = \begin{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} & \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} & \dots & \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \end{bmatrix}$

variations large





# Outline of section

- 1 Introduction
  - Motivation
  - Purpose of factor analysis
- 2 Modeling
  - The Orthogonal Factor Model
- 3 Method
  - Principal Component
  - MLE
- 4 Rotation
- 5 Factor Score**
- 6 Factor Analysis in R(Python)
- 7 Extended reading

# Factor Score

**Q:** What is factor Score?

**Ans:** the estimated values of the common factors. (not real values)

Suppose first that the mean vector  $\mu, \mathbf{L}, \Psi$  are known for the factor model

**Model:**  $\mathbf{X} - \mu = \mathbf{L} \underbrace{\mathbf{F}}_{\text{Random}} + \varepsilon$

**Data:** Further, regard the specific factors  $\varepsilon = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p]$  as errors. Since  $\text{Var}(\varepsilon_i) = \Psi_i, i = 1, 2, \dots, p$ , need not be equal.

**Solve:**

- WLSE:  $\hat{\mathbf{f}}_j = \mathbf{L}'\Psi^{-1}\mathbf{L}^{-1}\mathbf{L}'\Psi^{-1}(\mathbf{x} - \mu).$
- OLSE: (Principal component)  $\hat{\mathbf{f}}_j = \mathbf{L}'\mathbf{L}^{-1}\mathbf{L}'(\mathbf{x} - \mu)$
- Regression:  $\hat{\mathbf{f}} = \hat{\mathbf{L}}'\mathbf{S}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}}), j = 1, 2, \dots, n,$   
where  $\mathbf{S}$  (the original sample covariacne matrix)

# Outline of section

- 1 Introduction
  - Motivation
  - Purpose of factor analysis
- 2 Modeling
  - The Orthogonal Factor Model
- 3 Method
  - Principal Component
  - MLE
- 4 Rotation
- 5 Factor Score
- 6 Factor Analysis in R(Python)
- 7 Extended reading

# Example (Stock price data)

**Table 9.3**

Variable	Maximum likelihood			Principal components		
	Estimated factor loadings		Specific variances $\hat{\psi}_i = 1 - \hat{h}_i^2$	Estimated factor loadings		Specific variances $\tilde{\psi}_i = 1 - \tilde{h}_i^2$
	$F_1$	$F_2$		$F_1$	$F_2$	
1. J P Morgan	.115	.755	.42	.732	-.437	.27
2. Citibank	.322	.788	.27	.831	-.280	.23
3. Wells Fargo	.182	.652	.54	.726	-.374	.33
4. Royal Dutch Shell	1.000	-.000	.00	.605	.694	.15
5. Texaco	.683	-.032	.53	.563	.719	.17
Cumulative proportion of total (standardized) sample variance explained	.323	.647		.487	.769	

*Handwritten notes:*

- A bracket under the  $F_1$  and  $F_2$  loadings for the first five variables is labeled "not orthogonal".
- A bracket under the  $F_1$  and  $F_2$  loadings for the last five variables is labeled "orthogonal".

Python & R

# Outline of section

- 1 Introduction
  - Motivation
  - Purpose of factor analysis
- 2 Modeling
  - The Orthogonal Factor Model
- 3 Method
  - Principal Component
  - MLE
- 4 Rotation
- 5 Factor Score
- 6 Factor Analysis in R(Python)
- 7 Extended reading

- Mixed data type
  - R:FAMD - Factor Analysis of Mixed Data
  - Python factor analysis library (PCA, CA, MCA, MFA, FAMD)

Thank you for your kind attention.