Factor analysis

ZIH-SYUAN, JIAN

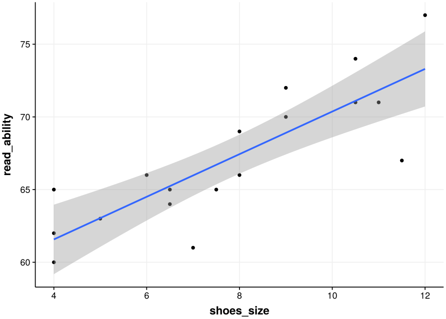
National Central University

September 24, 2018

Outline of section

Motivation example (Using a Regression example)

For children in elementrays school
Observed variable:Shoe size and reading ability



Purpose of factor analysis

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

Outline of section

The Orthogonal Factor Model

The observable random vector **X**, with components has mean μ and covariance matrix Σ. **X** is linearly dependent upon a few unobservable random variables $F_1, F_2, \ldots, F_m$ called common factor

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

in matrix notation

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

Assume that

$$
\begin{aligned}
\mathrm{E}(\mathbf{F}) = \mathbf{0}_{(m\times 1)}, \mathrm{Cov}(\mathbf{F}) = \mathrm{E}(\mathbf{FF}') = \mathrm{I}_{(m\times m)} \\
\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}_{(p\times 1)}, \mathrm{Cov}(\boldsymbol{\varepsilon}) = \mathrm{E}(\boldsymbol{\varepsilon\varepsilon}') = \Psi_{(p\times p)}
\end{aligned}
$$

where Ψ is a diagonal matrix.

So the covariance structure for **X** From this model

$$
\begin{aligned}
\boldsymbol{\Sigma} &= \mathsf{Cov}(\mathbf{X}) = \mathrm{E}\left((\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})'\right) \\
&= \mathbf{L}\mathrm{E}(\mathbf{F}\mathbf{F}')\mathbf{L}' + \cancelto{0}{\mathrm{E}(\boldsymbol{\varepsilon}\mathbf{F}')\mathbf{L}'} + \cancelto{0}{\mathbf{L}\mathrm{E}(\mathbf{F}\boldsymbol{\varepsilon}')} + \mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') \\
&= \mathbf{L}\mathbf{L}' + \Psi
\end{aligned}
$$

$\mathsf{Cov}(\varepsilon, \mathbf{F}) = \mathrm{E}(\varepsilon\mathbf{F}') = 0,$
$\mathsf{Cov}(\mathbf{X}, \mathbf{F}) = \mathrm{E}(\mathbf{X} - \boldsymbol{\mu}\mathbf{F}') = \mathbf{L}\mathrm{E}(\mathbf{F}\mathbf{F}') + \mathrm{E}\left(\boldsymbol{\varepsilon}\mathbf{F}'\right) = \mathbf{L}.$

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

Principlal Component

The spectral decompostion of provides is with one factoring of the covariance matrix Σ. Let Σ have eigenvalue-eigenvector pairs (λ_i, e_i) with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. Then

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

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

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

$$= \left[\sqrt{\lambda_1}e_1 \ \sqrt{\lambda_2}e_2 \ \cdots \ \sqrt{\lambda_m}e_m\right] \begin{bmatrix} \sqrt{\lambda_1}e_1' \\ \sqrt{\lambda_2}e_2' \\ \vdots \\ \sqrt{\lambda_p}e_m' \end{bmatrix}$$

$$+ \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_p \end{bmatrix}$$

Use **S** or **R** in place of $\boldsymbol{\Sigma}$

How to choose m?

The m is chosen such that a "suitable proportion" of the total sample variance is explained

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

Estimation by MLE

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

Outline of section

Rotation

- Factor rotation = Orthogonal transformation of the factors
- Different orthogonal transformation lead to the same estimated covariance matrix, residual matrix, specific variances, and communalities
  $\mathbf{L}^* = \mathbf{LT}$ then $\mathbf{LL}' + \Psi = \mathbf{LTT'L'} + \Psi = \mathbf{L^*L^{*\prime}} + \Psi$
- 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 $\tilde{\ell}^*_{ij} = \hat{\ell}^*_{ij}/\hat{h}_i$ 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

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

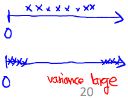
as large as possible.

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

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

* Idea: $L = \left[\ \cdots\ \right]_{p\times m} \longrightarrow L^* = \left[\ \cdots\ \right]$

variations large



Outline of section

Factor Score

Q: What is factor Score?

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

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

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

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

Solve:

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

Outline of section

Example (Stock price data)

Table 9.3

Variable	Maximum likelihood: Estimated factor loadings F_1	F_2	Specific variances $\hat{\psi}_i = 1 - \hat{h}_i^2$	Principal components: Estimated factor loadings F_1	F_2	Specific variances $\tilde{\psi}_i = 1 - \tilde{h}_i^2$
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	

not orthogonal

orthogonal

Python & R

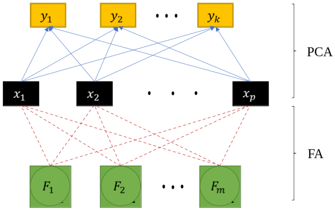
Outline of section

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

Outline of section

Conclusion

- Principal component analysis (PCA) is to summarize a set of variables with a few components by using their linear combinations.
- Factor analysis (FA) is to find latent variables behind a set of variables.
- Both methods deal with the covariance structure of variables
  - PCA focuses on the variances
  - FA focuses on the correlations



Thank you for your kind attention.