—WORKSHOP—

Applied Classical and Modern Multivariate Statistical Analysis

Module 6: Factor Analysis

Weixing Song, Juan Du

Department of Statistics
Kansas State University

Outline

Outline

What is Factor Analysis?

The purpose of factor analysis is to describe the covariance relationship among many variables in terms of a few underlying, but unobservable, random quantities called factors.

Motivation: Suppose all the variables within a particular group are highly correlated among themselves, but have relatively small correlations with variables in a different group. Each group of variables represents a single underlying factor.

Example: Test scores (Spearman(1904))

$$\begin{bmatrix} & Classics & French & Eng & Math & Discr & Music \\ Classics & & 0.83 & 0.78 & 0.7 & 0.66 & 0.63 \\ French & & & 0.67 & 0.67 & 0.65 & 0.57 \\ Eng & & & & 0.64 & 0.54 & 0.51 \\ Math & & & & & 0.45 & 0.51 \\ Discr & & & & & & 0.40 \\ Music & & & & & & \end{bmatrix}$$

Outline

Random vector $\mathbf{X}$: $E\mathbf{X} = \boldsymbol{\mu}$, $\text{Cov}(\mathbf{X}) = \Sigma$.

The factor model postulates that $\mathbf{X}$ is linearly dependent upon a few unobservable random variables $F_1, \ldots, F_m$, called common factors, and p additional sources of variation $\varepsilon_1, \ldots, \varepsilon_p$, called errors/specific factors.

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

Matrix form of the factor model

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

The coefficient l_{ij} is called the loading of the i-th variable on the j-th factor, and $\mathbf{L}$ is called loading matrix.

Orthogonal Factor Model with m Common Factors

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

where

- μ_i = mean of variable i;
- ε_i = i-th specific factor;
- F_j = j-th common factor;
- l_{ij} =loading of the i-th variable on the j-th factor.

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

- $\mathbf{F}$ and $\boldsymbol{\varepsilon}$ are independent;
- $E(\mathbf{F}) = 0$, $\mathrm{Cov}(\mathbf{F}) = \mathbf{I}$;
- $E(\boldsymbol{\varepsilon}) = 0$, $\mathrm{Cov}(\boldsymbol{\varepsilon}) = \Psi$, where Ψ is a diagonal matrix.

Covariance Structure for the Orthogonal Factor Model

- $\mathrm{Cov}(\mathbf{X}) = \mathbf{L}\mathbf{L}' + \Psi$ or

$$\mathrm{Var}(X_i) = \sigma_{ii} = l_{i1}^2 + \cdots + l_{im}^2 + \psi_i = \underset{\text{communality}}{{h_i}^2} + \underset{\text{specific variance}}{\psi_{\mathrm{i}}}$$

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

- $\mathrm{Cov}(\mathbf{X}, \mathbf{F}) = \mathbf{L}$, or $\mathrm{Cov}(X_i, F_j) = l_{ij}$

Remarks:

- The factor model assumes that the $p(p+1)/2$ variances and covariances for $\mathbf{X}$ can be reproduced from the pm factor loadings l_{ij} and the p specific variances ψ_i.
- It is when m is small relative to p that factor analysis is most useful.
- Even the covariance matrix can be factored as $\mathbf{LL}' + \Psi$, the loading matrix $\mathbf{L}$ and the factors $\mathbf{F}$ are ambiguous (not unique).

Ambiguity of Factor Analysis

Factor loadings $\mathbf{L}$ are determined only up to an orthogonal matrix $\mathbf{T}$. Thus, the loadings

$$\mathbf{L}^* = \mathbf{LT} \quad \text{and} \quad \mathbf{T}$$

both give the same representation. The communalities, given by the diagonal elements of $\mathbf{LL}' = \mathbf{L}^*\mathbf{L}^{*'}$ are also unaffected by the choice of $\mathbf{T}$.

Outline

The principal component method is based on the spectral decomposition of Σ.

Let $(\lambda_i, \mathbf{e}_i)$ with $\lambda_1 \geq \cdots \geq \lambda_p$ be the eigenvalue-eigenvector pair of Σ. The spectral decomposition of Σ is given by

$$\begin{aligned}\Sigma &= \lambda_1 \mathbf{e}_1 \mathbf{e}_1' + \lambda_2 \mathbf{e}_2 \mathbf{e}_2' + \cdots + \lambda_p \mathbf{e}_p \mathbf{e}_p' \\ &= \left[\sqrt{\lambda_1}\mathbf{e}_1,\ \sqrt{\lambda_2}\mathbf{e}_2,\ \cdots,\ \sqrt{\lambda_p}\mathbf{e}_p\right] \begin{bmatrix} \sqrt{\lambda_1}\mathbf{e}_1' \\ \sqrt{\lambda_2}\mathbf{e}_2' \\ \vdots \\ \sqrt{\lambda_p}\mathbf{e}_p' \end{bmatrix}\end{aligned}$$

Suppose $\lambda_{m+1}, \ldots, \lambda_p$ are "small", then

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

where $\psi_i = \sigma_{ii} - \sum_{j=1}^{m} l_{ij}^2$.

The principal component factor analysis of the sample covariance matrix $\mathbf{S}$ is specified in terms of its eigenvalue-eigenvector pairs $(\hat{\lambda}_1, \hat{\mathbf{e}}_1)$, ..., $(\hat{\lambda}_p, \hat{\mathbf{e}}_p)$, 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 $\{\tilde{l}_{ij}\}$ is given by

$$\tilde{\mathbf{L}} = \left[\sqrt{\hat{\lambda}_1}\hat{\mathbf{e}}_1, \cdots, \sqrt{\hat{\lambda}_m}\hat{\mathbf{e}}_m\right]$$

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

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

Communities are estimated as

$$\tilde{h}_i^2 = \sum_{j=1}^{m} \tilde{l}_{ij}^2$$

The principal component factor analysis of the sample correlation matrix is obtained by starting with $\mathbf{R}$ in place of $\mathbf{S}$.

How many common factors?

Analytically, we have

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

The contributions of the first few factors to the sample variances of the variables should be large.

The contribution to the sample variance s_{ii} from the first common factor is $\tilde{l}_{i1}^2$, and the contribution to the total sample variance $s_{11} + \cdots + s_{pp}$ from the first common factor is then

$$\tilde{l}_{11}^2 + \tilde{l}_{21}^2 + \cdots + \tilde{l}_{p1}^2 = (\sqrt{\hat{\lambda}_1}\hat{\mathbf{e}}_1)'(\sqrt{\hat{\lambda}_1}\hat{\mathbf{e}}_1) = \hat{\lambda}_1.$$

One can select the number of common factors based on the

$$\begin{array}{c}\text{Proportion of total sample}\\ \text{variance due to} j\text{-th factor}\end{array} = \begin{cases} \hat{\lambda}_j/(s_{11} + \cdots + s_{pp}) & \text{for a factor analysis of } \mathbf{S}, \\ \hat{\lambda}_j/p & \text{for a factor analysis of } \mathbf{R}. \end{cases}$$

Remark: In some statistics software, the value of m is set to be the number of eigenvalues of $\mathbf{R}$ greater than 1 if the sample correlation matrix is factored.

Example 6.1 Factor analysis of consumer-preference data
In a consumer-preference study, a random sample of customers were asked to rate several attributes of a new product. The responses, on a 7-point semantic differential scale, were tabulated and the attribute correlation matrix constructed. The correlation matrix is presented next:

Attribute (Variable)		1	2	3	4	5
Taste	1	1.00	.02	.96	.42	.01
Good buy for money	2	.02	1.00	.13	.71	.85
Flavor	3	.96	.13	1.00	.50	.11
Suitable for snack	4	.42	.71	.50	1.00	.79
Provides lots of energy	5	.01	.85	.11	.79	1.00

(See R-code and output)

Principal Component Solution

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		

Factorization

$$\widetilde{\mathbf{L}}\widetilde{\mathbf{L}}'+\widetilde{\boldsymbol{\Psi}}=\begin{bmatrix}0.56 & -0.82\\0.78 & 0.53\\0.65 & -0.75\\0.94 & 0.10\\0.80 & 0.54\end{bmatrix}\begin{bmatrix}0.56 & 0.78 & 0.65 & 0.94 & 0.80\\-0.82 & 0.53 & -0.75 & 0.10 & 0.54\end{bmatrix}$$

$$+\begin{bmatrix}0.02 & 0 & 0 & 0 & 0\\0 & 0.12 & 0 & 0 & 0\\0 & 0 & 0.02 & 0 & 0\\0 & 0 & 0 & 0.11 & 0\\0 & 0 & 0 & 0 & 0.07\end{bmatrix}=\begin{bmatrix}1 & 0.01 & 0.97 & 0.44 & 0.00\\ & 1 & 0.11 & 0.79 & 0.91\\ & & 1 & 0.53 & 0.11\\ & & & 1 & 0.81\\ & & & & 1\end{bmatrix}$$

Assume that $\mathbf{F}$ and ε follow normal distributions. The likelihood function is given by

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

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

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

where $\Sigma = \mathbf{L}\mathbf{L}' + \Psi$.

We maximize $L(\boldsymbol{\mu}, \Sigma)$ subject to the uniqueness condition $\mathbf{L}'\Psi^{-1}\mathbf{L}$ to be diagonal.

The maximum likelihood estimates of the communalities are

$$\hat{h}_i^2 = \hat{\ell}_{i1}^2 + \hat{\ell}_{i2}^2 + \cdots + \hat{\ell}_{im}^2 \quad \text{for } i = 1, 2, \ldots, p \tag{9-27}$$

so

$$\begin{pmatrix} \text{Proportion of total sample} \\ \text{variance due to } j\text{th factor} \end{pmatrix} = \frac{\hat{\ell}_{1j}^2 + \hat{\ell}_{2j}^2 + \cdots + \hat{\ell}_{pj}^2}{s_{11} + s_{22} + \cdots + s_{pp}} \tag{9-28}$$

Example 6.2 The weekly rates of return for five stocks (JP Morgan, Citibank, Wells Fargo, Royal Dutch Shell, and ExxonMobil) listed on the New York Stock Exchange were determined for the period January 2004 through December 2005. The weekly rates of return are defined as (current Friday closing price - previous Friday closing price)/(previous Friday closing price), adjusted for stock splits and dividends. The observations in 103 successive weeks appear to be independently distributed, but the rates of return across stocks are correlated, since, as one expects, stocks tend to move together in response to general economic conditions. Let x_1, x_2, x_3, x_4, x_5 denote the observed weekly rates of return for JP Morgan, Citibank, Wells Fargo, Royal Dutch Shell, and ExxonMobil, respectively. Find the PCs based on the sample correlation coefficient. (See R-code and output)

Outline

Factor Rotation

Factor Rotation: $\hat{\mathbf{L}}^* = \hat{\mathbf{L}}\mathbf{T}$ where $\mathbf{T}\mathbf{T}' = \mathbf{I}, \mathbf{T}'\mathbf{T} = \mathbf{I}$

Varimax Criterion

Let $\tilde{l}^*_{ij} = \hat{l}^*_{ij}/\hat{h}_i$ and select $\mathbf{T}$ that makes

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

as large as possible.

Interpretation: Note that

$$V \propto \sum_{j=1}^{m}(\text{varinace of squares of scaled loadings for } j\text{-th factor}),$$

so such a $\mathbf{T}$ can "spread out" the squares of the loadings on each factor as much as possible.

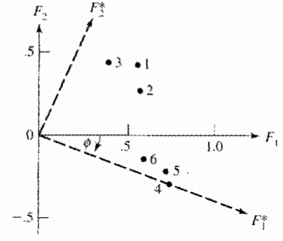
Example 6.3 (Examination Scores)Lawley and Maxwell(1971)

$$
\begin{array}{c c}
 & \begin{array}{cccccc} \text{Gaelic} & \text{English} & \text{History} & \text{Arithmetic} & \text{Algebra} & \text{Geometry} \end{array} \\
\mathbf{R} = & \begin{bmatrix}
1.0 & .439 & .410 & .288 & .329 & .248 \\
 & 1.0 & .351 & .354 & .320 & .329 \\
 & & 1.0 & .164 & .190 & .181 \\
 & & & 1.0 & .595 & .470 \\
 & & & & 1.0 & .464 \\
 & & & & & 1.0
\end{bmatrix}
\end{array}
$$

Maximum Likelihood Solution

Variable	Estimated factor loadings F_1	F_2	Communalities $\hat{h}_i^2$
1. Gaelic	.553	.429	.490
2. English	.568	.288	.406
3. History	.392	.450	.356
4. Arithmetic	.740	−.273	.623
5. Algebra	.724	−.211	.569
6. Geometry	.595	−.132	.372

Example 6.3 (Continued) Factor Rotation



Rotated Factor Loading

Variable	Estimated rotated factor loadings F_1^*	F_2^*	Communalities $\hat{h}_i^{*2} = \hat{h}_i^2$
1. Gaelic	.235	.659	.490
2. English	.323	.549	.406
3. History	.088	.590	.356
4. Arithmetic	.771	.170	.623
5. Algebra	.724	.213	.569
6. Geometry	.572	.210	.372

Example 6.2 (Revisited. Factor rotation using R)

The weekly rates of return for five stocks (JP Morgan, Citibank, Wells Fargo, Royal Dutch Shell, and ExxonMobil) listed on the New York Stock Exchange were determined for the period January 2004 through December 2005. The weekly rates of return are defined as (current Friday closing price - previous Friday closing price)/(previous Friday closing price), adjusted for stock splits and dividends. The observations in 103 successive weeks appear to be independently distributed, but the rates of return across stocks are correlated, since, as one expects, stocks tend to move together in response to general economic conditions.
Let x_1, x_2, x_3, x_4, x_5 denote the observed weekly rates of return for JP Morgan, Citibank, Wells Fargo, Royal Dutch Shell, and ExxonMobil, respectively. Find the PCs based on the sample correlation coefficient. (See R-code and output)