Factor Analysis

- What is factor analysis?
- Orthogonal Factor Model
- Methods of estimation
- Principal component estimation
- Factor rotation

What is Factor Analysis?

- A statistics procedure that explores the covariance structure among a number of observable variables which are conceptualized as manifestation of a few underlying unobservable random quantities called factors.

- According to the above conceptual framework, formulate a factor model in order to capture the relations between the factors and the original variables and suggest meaningful interpretations to understand the variability of the observable data.

Orthogonal Factor Model

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

In matrix notation,

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

Where $\boldsymbol{L}$ is the $p \times m$ matrix of factor loadings l_{ij}

<u>Assumptions</u>

1. The common factors ($\boldsymbol{F}$) are independent of the specific factors ($\boldsymbol{\varepsilon}$). That is, $Cov(\boldsymbol{\varepsilon}, \boldsymbol{F}) = E(\boldsymbol{\varepsilon}\boldsymbol{F}') = \mathbf{0}_{(p \times m)}$
2. $E(\boldsymbol{F}) = \mathbf{0}, Cov(\boldsymbol{F}) = \boldsymbol{I}$
3. $E(\boldsymbol{\varepsilon}) = \mathbf{0}, Cov(\boldsymbol{\varepsilon}) = \boldsymbol{\Psi}$ ($\boldsymbol{\Psi}$ is a diagonal matrix with elements $\psi_1, \psi_2, \ldots, \psi_p$.)

Covariance structure of the Orthogonal Factor Model

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

   a. $\sigma_{ii} = Var(X_i) = l_{i1}^2 + \cdots + l_{im}^2 + \psi_i$

      Let $h_i^2 = l_{i1}^2 + \cdots + l_{im}^2 =$ communality, $\psi_i =$ specific variance
      Then $\sigma_{ii} = h_i^2 + \psi_i, \quad i = 1,2,\ldots,p$

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

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

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

Example

$$Cov(\boldsymbol{X}) = \boldsymbol{\Sigma} = \begin{bmatrix} \mathbf{19} & \mathbf{30} & \mathbf{2} & \mathbf{12} \\ \mathbf{30} & \mathbf{57} & \mathbf{5} & \mathbf{23} \\ \mathbf{2} & \mathbf{5} & \mathbf{38} & \mathbf{47} \\ \mathbf{12} & \mathbf{23} & \mathbf{47} & \mathbf{68} \end{bmatrix}$$

Example

$$Cov(\boldsymbol{X}) = \boldsymbol{\Sigma} = \begin{bmatrix} \mathbf{1} & \mathbf{.9} & \mathbf{.7} \\ \mathbf{.9} & \mathbf{1} & \mathbf{.4} \\ \mathbf{.7} & \mathbf{.4} & \mathbf{1} \end{bmatrix}$$

Method of Estimation

1. The Principal Component Method
   - By spectral decomposition of the sample covariance matrix $\boldsymbol{S}$ or the sample correlation matrix $\boldsymbol{R}$

2. The Maximum Likelihood Method
   - The common factors F and the specific factors $\boldsymbol{\varepsilon}$ are assumed to be normally distributed
   - The likelihood function can then be derived and numerical methods is used to compute the MLE

The Principal Component Method

Let the sample covariance matrix be $\boldsymbol{S}$

Let the eigenvalue-eigenvector pairs of $\boldsymbol{S}$ be $(\hat{\lambda}_1, \hat{\boldsymbol{e}}_1), (\hat{\lambda}_2, \hat{\boldsymbol{e}}_2), \cdots (\hat{\lambda}_p, \hat{\boldsymbol{e}}_p)$ where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_p$. Assume that there are $m < p$ number of common factors. The matrix of estimated factor loadings $\{\hat{l}_{ij}\}$ is given by

$$\hat{\boldsymbol{L}} = \left[\sqrt{\hat{\lambda}_1}\hat{\boldsymbol{e}}_1 \middle| \sqrt{\hat{\lambda}_2}\hat{\boldsymbol{e}}_2 \middle| \cdots\cdots\cdots \middle| \sqrt{\hat{\lambda}_m}\hat{\boldsymbol{e}}_m \right]$$

The Principal Component Method

The estimated specific variances are given by the diagonal elements of $\boldsymbol{S} - \hat{\boldsymbol{L}}\,\hat{\boldsymbol{L}}'$, Hence

$$\hat{\boldsymbol{\Psi}} = \begin{bmatrix} \hat{\psi}_1 & 0 & \cdots & 0 \\ 0 & \hat{\psi}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{\psi}_p \end{bmatrix}$$

with $\hat{\psi}_i = s_{ii} - \sum_{j=1}^{m} \hat{l}_{ij}^2$. Communalities are estimated as

$$\hat{h}_i^2 = \hat{l}_{i1}^2 + \cdots + \hat{l}_{im}^2$$

The principal component factor analysis of the sample correlation matrix is computed based on $\boldsymbol{R}$ instead of $\boldsymbol{S}$.

Factor rotation

1. Factor rotation: orthogonal transformation of the factor loadings, and the implied orthogonal transformation of the factors.
2. Rotation for simpler structure and hence facilitate interpretation.
3. The task is related to the attempt rotate the loadings such that each variable loads highly on a single factor and has relatively small loadings on the remaining factors.

Factor Rotation

Let $\boldsymbol{T}$ be an $m \times m$ orthogonal matrix. Therefore $\boldsymbol{TT}' = \boldsymbol{I}$

The equation on p.3

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

can be rewritten as

$$\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{LTT}'\boldsymbol{F} + \boldsymbol{\varepsilon} = \boldsymbol{\mu} + \boldsymbol{L}^*\boldsymbol{F}^* + \boldsymbol{\varepsilon}$$

where $\boldsymbol{L}^* = \boldsymbol{LT}$, $\boldsymbol{F}^* = \boldsymbol{T}'\boldsymbol{F}$. Note that

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

$$Cov(\boldsymbol{F}^*) = \boldsymbol{T}'\boldsymbol{Cov}(\boldsymbol{F})\boldsymbol{T} = \boldsymbol{T}'\boldsymbol{T} = \boldsymbol{I}$$

Varimax (Kaiser)

Let $\hat{l}_{ij}^{*}$ be the factor loading after rotation. Define $\tilde{l}_{ij}^{*} = \hat{l}_{ij}^{*}/\hat{h}_i$ to be the final rotated coefficients scaled by the square root of the communalities.

Varimax: Selects the orthogonal transformation T such that

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

is maximized.