Lecture 9 Factor Analysis

Shiwei Lan[1]

[1]School of Mathematical and Statistical Sciences
Arizona State University

STP533 Multivariate Analysis
Spring 2025

Table of Contents

Factor Analysis

- The modern beginnings lie in the early-20th-century attempts of Karl Pearson, Charles Spearman and others to define and measure intelligence.
- The **essential purpose** of factor analysis is to describe, if possible, the covariance relationships among many variables in terms of a few underlying, but unobservable, random quantities called *factors*.
- The factor model is motivated by the argument that variables can be grouped by their correlations: all variables with high correlations should be grouped together while having relatively small correlations with other groups.
- Factor analysis can be considered as an extension of principal component analysis (PCA). The approximation to the covariance matrix is more elaborate.

Orthogonal Factor Model

- Assume **observable** p-dimensional random vector $\mathbf{X} \sim (\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
- The factor model postulates that $\mathbf{X}$ is linearly dependent on a few **unobservable** random variables called *factors*, $\mathbf{F} = [F1, \cdots, F_m]$ and p additional source variation called *errors*, $\boldsymbol{\epsilon} = [\epsilon_1, \cdots, \epsilon_p]$:

$$\mathbf{X}_{p\times 1} - \boldsymbol{\mu} = \mathbf{L}_{p\times m}\mathbf{F}_{m\times 1} + \boldsymbol{\epsilon}_{p\times 1}$$

- The coefficient $\mathbf{L}$ is called the *matrix of factor loadings*.
- Additionally, we assume

$$\mathrm{E}(\mathbf{F}) = \mathbf{0}_{m\times 1}, \quad \mathrm{Cov}(\mathbf{F}) = \mathbf{I}_m; \quad \mathrm{E}(\boldsymbol{\epsilon}) = \mathbf{0}_{p\times 1}, \quad \mathrm{Cov}(\boldsymbol{\epsilon}) = \boldsymbol{\Psi}_{p\times p} = \mathrm{diag}(\boldsymbol{\psi})$$

Orthogonal Factor Model

- Now let us investigate the covariance matrix of $\mathbf{X}$:

$$\mathrm{Cov}(\mathbf{X}) = \mathrm{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T]$$

- How about $\mathrm{Cov}(\mathbf{X}, \mathbf{F})$?

Orthogonal Factor Model

- In fact, we have $\mathrm{Cov}(\mathbf{X}) = \mathbf{L}\mathbf{L}^T + \mathbf{\Psi}$, or equivalently

$$\mathrm{Var}(X_i) = \sigma_{ii} = \underbrace{h_i^2}_{\textit{communality}} + \underbrace{\psi_i}_{\textit{specific variance}}, \quad h_i^2 = \sum_{j=1}^{m} \ell_{ij}^2$$

$$\mathrm{Cov}(x_i, k) = \sum_{j=1}^{m} \ell_{ij}\ell_{kj}$$

- And $\mathrm{Cov}(\mathbf{X}, \mathbf{F}) = \mathbf{L}$ or $\mathrm{Cov}(X_i, F_j) = \ell_{ij}$.
- Note that $\mathbf{X} - \boldsymbol{\mu} = \mathbf{L}\mathbf{F} + \boldsymbol{\epsilon}$ is linear in the common factors $\mathbf{F}$.

Orthogonal Factor Model

- The factor model assumes that the $p + \binom{p}{2} = p(p+1)/2$ variances and covariances of $\mathbf{X}$ can be reproduced from the pm factor loadings ℓ_{ij} and the p specific variances ψ_i.
- When $m = p$, any covariance matrix $\mathbf{\Sigma}$ can be reproduced exactly as $\mathbf{L}\mathbf{L}^T$ so we can set $\mathbf{\Psi} = \mathbf{0}$.
- However, when $m < p$ the factor analysis is most useful. Unfortunately, most covariance matrices cannot be factored as $\mathbf{L}\mathbf{L}^T + \mathbf{\Psi}$ when $m \ll p$.

Example 9.2 (Nonexistence of a proper solution) Let $p = 3$ and $m = 1$, and suppose the random variables X_1, X_2, and X_3 have the positive definite covariance matrix

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & .9 & .7 \\ .9 & 1 & .4 \\ .7 & .4 & 1 \end{bmatrix}$$

Using the factor model in (9-4), we obtain

$$\begin{aligned} X_1 - \mu_1 &= \ell_{11}F_1 + \varepsilon_1 \\ X_2 - \mu_2 &= \ell_{21}F_1 + \varepsilon_2 \\ X_3 - \mu_3 &= \ell_{31}F_1 + \varepsilon_3 \end{aligned}$$

The covariance structure in (9-5) implies that

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

Inherent Ambiguity

- Let $\mathbf{T}$ be any $m \times m$ orthogonal matrix, i.e. $\mathbf{T}\mathbf{T}^T = \mathbf{T}^T\mathbf{T} = \mathbf{I}_m$.
- Then we can re-write the factor model as

$$\mathbf{X} - \boldsymbol{\mu} = \mathbf{L}\mathbf{F} + \boldsymbol{\epsilon} = \mathbf{L}^*\mathbf{F}^* + \boldsymbol{\epsilon}, \quad \mathbf{L}^* = \mathbf{L}\mathbf{T}, \quad \mathbf{F}^* = \mathbf{T}^T\mathbf{F}$$

- We also have (why?)

$$\mathrm{E}(\mathbf{F}^*) = \mathbf{0}, \quad \mathrm{Cov}(\mathbf{F}^*) = \mathbf{I}_m$$

- And finally

$$\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^T + \boldsymbol{\Psi} = \mathbf{L}^*(\mathbf{L}^*)^T + \boldsymbol{\Psi}$$

- Therefore, the factor loadings $\mathbf{L}$ can be determined only up to an orthogonal matrix $\mathbf{T}$, i.e. $\mathbf{L}^* = \mathbf{L}\mathbf{T}$ and $\mathbf{L}$ yield the same factor model.
- Then how to determine $\mathbf{L}$?

Table of Contents

Methods of Estimation

- Let $\mathbf{X}_{n\times p}$ be a sample of size n with population mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$.
- How to find $\mathbf{F}_{n\times m}$ and $\mathbf{L}_{p\times m}$ such that $\mathbf{X}_{n\times p} = \boldsymbol{\mu}_{n\times p} + \mathbf{F}_{n\times m}\mathbf{L}^T_{m\times p} + \mathbf{E}_{n\times p}$?
- While $\boldsymbol{\Sigma}$ is unknown, we can estimate it with sample covariance $\mathbf{S}$.
- If the off-diagonal elements of $\mathbf{S}$ are small, the variables appears not related, and a factor analysis is not very useful.
- However, when Σ appears to deviate significantly from a diagonal matrix, then a factor model can be helpful.
- We will consider two methods of estimation: the principal componential (PC) method and the maximum likelihood (ML) method.

The Principal Component Method

- A natural candidate is from the principal component analysis (PCA).
- Consider the spectral decomposition of $\mathbf{\Sigma}$

$$\mathbf{\Sigma}_{p\times p} = \mathbf{\Gamma\Lambda\Gamma}^T = \mathbf{LL}^T + \mathbf{\Psi}, \quad \mathbf{L}_{p\times p} = \mathbf{\Gamma\Lambda}^{\frac{1}{2}}, \quad \mathbf{\Psi} = \mathbf{0}$$

- Now if we consider the m PCs in $\mathbf{L}$:

$$\mathbf{\Sigma}_{p\times p} \approx \mathbf{LL}^T + \mathbf{\Psi}, \quad \mathbf{L}_{p\times m} = \mathbf{\Gamma}_{p\times m}\mathbf{\Lambda}^{\frac{1}{2}}_{m\times m}, \ \mathbf{\Psi} = \mathrm{diag}(\{\psi_i\}), \ \psi_i = \sigma_{ii} - \sum_{j=1}^m \ell_{ij}^2$$

- In practice, we consider standardized variables with sample mean $\bar{\mathbf{X}}$ and sample covariance $\mathbf{S}$, i.e. $\mathbf{Z} = (\mathbf{X} - \bar{\mathbf{X}})\mathrm{diag}(\mathbf{S})^{-\frac{1}{2}}$.

Principal Component Solution

- Now we consider the sample PCs of the sample covariance $\mathbf{S} = \hat{\mathbf{\Gamma}}\hat{\mathbf{\Lambda}}\hat{\mathbf{\Gamma}}^T$.
- Let $\hat{\mathbf{L}}$ be the m sample PCs. Then we have

$$\mathbf{S}_{p\times p} \approx \hat{\mathbf{L}}\hat{\mathbf{L}}^T + \hat{\mathbf{\Psi}}, \quad \hat{\mathbf{L}}_{p\times m} = \hat{\mathbf{\Gamma}}_{p\times m}\hat{\mathbf{\Lambda}}^{\frac{1}{2}}_{m\times m}, \; \hat{\mathbf{\Psi}} = \mathrm{diag}(\{\hat{\psi}_i\}), \; \hat{\psi}_i = s_{ii} - \sum_{j=1}^m \hat{\ell}_{ij}^2$$

- In practice, the factor analysis can also be done for the sample correlation matrix $\boldsymbol{P}$.
- The number of common factors, m, can be determined similarly as in PCA by thresholding the proportion of total sample variance: $\hat{\lambda}_m/\mathrm{tr}(\mathbf{S})$ or $\hat{\lambda}_m/p$.

Factor Analysis of Consumer-Preference Data

Example 9.3 (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

It is clear from the circled entries in the correlation matrix that variables 1 and 3 and variables 2 and 5 form groups. Variable 4 is "closer" to the (2, 5) group than the (1, 3) group. Given these results and the small number of variables, we might expect that the apparent linear relationships between the variables can be explained in terms of, at most, two or three common factors.

Factor Analysis of Stock-Price Data

Example 9.4 (Factor analysis of stock-price data) Stock-price data consisting of $n = 103$ weekly rates of return on $p = 5$ stocks were introduced in Example 8.5. In that example, the first two sample principal components were obtained from $\mathbf{R}$. Taking $m = 1$ and $m = 2$, we can easily obtain principal component solutions to the orthogonal factor model. Specifically, the estimated factor loadings are the sample principal component coefficients (eigenvectors of $\mathbf{R}$), scaled by the square root of the corresponding eigenvalues. The estimated factor loadings, communalities, specific variances, and proportion of total (standardized) sample variance explained by each factor for the $m = 1$ and $m = 2$ factor solutions are available in Table 9.2. The communalities are given by (9-17). So, for example, with $m = 2$, $\tilde{h}_1^2 = \tilde{\ell}_{11}^2 + \tilde{\ell}_{12}^2 = (.732)^2 + (-.437)^2 = .73$.

A Modified Approach*: the Principal Factor Solution

- Now consider the factor analysis in terms of the sample correlation $\boldsymbol{P}$.
- If the factor model $\boldsymbol{\rho} = \mathbf{L}\mathbf{L}^T + \boldsymbol{\Psi}$ is correctly specified, we have

$$\rho_{ii} = 1 = h_i^2 + \psi_i, \quad i.e.\ \boldsymbol{\rho} - \Psi = \mathbf{L}\mathbf{L}^T$$

- Suppose the initial estimates ψ_i^* of the specific variances are available. We obtain a "reduced" sample correlation matrix $\boldsymbol{P}_r$ by replacing the diagonal elements (1) with $h_i^{*2} = 1 - \psi_i^*$.
- If $\boldsymbol{P}_r = \hat{\boldsymbol{\Gamma}}_r \hat{\boldsymbol{\Lambda}}_r \hat{\boldsymbol{\Gamma}}_r^T$ is factored as

$$\boldsymbol{P}_r \approx \mathbf{L}_r^* \mathbf{L}_r^{*T}, \quad (\mathbf{L}_r^*)_{p\times m} = (\hat{\boldsymbol{\Gamma}}_r)_{p\times m} (\hat{\boldsymbol{\Lambda}}_r^{\frac{1}{2}})_{m\times m}$$

- Then principal factor method yields

$$\psi_i^* = 1 - \sum_{j=1}^m \ell_{ij}^{*2}, \quad \hat{h}_i^{*2} = \sum_{j=1}^m \ell_{ij}^{*2}$$

- The solution can be constructed iteratively with the communality estimates $\hat{h}_i^{*2}$ being the initial estimates for the next stage.

The Maximum Likelihood Method

- If the common factors $\mathbf{F}$ and the specific factors ϵ are assumed to be normally distributed, then the maximum likelihood estimates (MLE) can be obtained.
- Suppose

$$\mathbf{F} \sim N_m(\mathbf{0}, \mathbf{I}_m), \quad \epsilon \sim N_p(\mathbf{0}, \mathbf{\Psi})$$

- Then we have $\mathbf{X} = \boldsymbol{\mu} + \mathbf{LF} + \epsilon \sim N_p(\boldsymbol{\mu}, \mathbf{\Sigma})$ with $\mathbf{\Sigma} = \mathbf{LL}^T + \mathbf{\Psi}$.
- The likelihood of sample $\{\mathbf{x}_i\}_{i=1}^n$ becomes

$$\begin{aligned} L(\boldsymbol{\mu}, \mathbf{\Sigma}) =& (2\pi)^{-\frac{np}{2}} |\mathbf{\Sigma}|^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})^T \mathbf{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \right\} \\ =& (2\pi)^{-\frac{(n-1)p}{2}} |\mathbf{\Sigma}|^{-\frac{n-1}{2}} \exp\left\{ -\frac{n-1}{2} \mathrm{tr}\left[\mathbf{\Sigma}^{-1} \mathbf{S} \right] \right\} \cdot \\ & (2\pi)^{-\frac{p}{2}} |\mathbf{\Sigma}|^{-\frac{1}{2}} \exp\left\{ -\frac{n}{2} (\bar{\mathbf{x}} - \boldsymbol{\mu})^T \mathbf{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \right\} \end{aligned}$$

The Maximum Likelihood Method

- To obtain the unique solution, we need the following condition

$$\mathbf{L}^T \mathbf{\Psi} \mathbf{L} = \Delta \textit{ is diagonal}$$

- In general, we numerically optimize the log-likelihood to obtain $\hat{\mathbf{L}}$ and $\hat{\mathbf{\Psi}}$.
- In the meantime, it is straightforward to see that $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$.
- The MLE of the communalities are $\hat{h}_i^2 = \sum_{j=1}^m \hat{\ell}_{ij}^2$.
- And the proportion of the total sample variance due to the j-th factor is $\frac{\sum_{i=1}^p \hat{\ell}_{ij}^2}{\mathrm{tr}(\mathbf{S})}$.

The Maximum Likelihood Method

- For standardized variables $\mathbf{Z} = \mathbf{V}^{-\frac{1}{2}}(\mathbf{X} - \boldsymbol{\mu})$, the covariance of $\mathbf{Z}$ can be factorized as

$$\boldsymbol{\rho} = \mathbf{V}^{-\frac{1}{2}}\boldsymbol{\Sigma}\mathbf{V}^{-\frac{1}{2}} = \mathbf{L}_\mathbf{z}\mathbf{L}_\mathbf{z}^T + \boldsymbol{\Psi}_\mathbf{z}, \quad \mathbf{L}_\mathbf{z} = \mathbf{V}^{-\frac{1}{2}}\mathbf{L}, \; \boldsymbol{\Psi}_\mathbf{z} = \mathbf{V}^{-\frac{1}{2}}\boldsymbol{\Psi}\mathbf{V}^{-\frac{1}{2}}$$

- MLE for $\boldsymbol{\rho}$ can be obtained based on that for factors of $\boldsymbol{\Sigma}$:

$$\hat{\boldsymbol{\rho}} = \hat{\mathbf{L}}_\mathbf{z}\hat{\mathbf{L}}_\mathbf{z}^T + \hat{\boldsymbol{\Psi}}_\mathbf{z}, \quad \hat{\mathbf{L}}_\mathbf{z} = \hat{\mathbf{V}}^{-\frac{1}{2}}\hat{\mathbf{L}}, \; \hat{\boldsymbol{\Psi}}_\mathbf{z} = \hat{\mathbf{V}}^{-\frac{1}{2}}\hat{\boldsymbol{\Psi}}\hat{\mathbf{V}}^{-\frac{1}{2}}$$

  where $\hat{\mathbf{V}}$ is the MLE of $\mathbf{V}$.
- The MLE of the communalities for $\boldsymbol{\rho}$ are $\hat{h}^2_{\mathbf{z},i} = \sum_{j=1}^m \hat{\ell}^2_{\mathbf{z},ij}$. The proportion of the total (standardized) sample variance due to the j-th factor is $\frac{\sum_{i=1}^p \hat{\ell}^2_{\mathbf{z},ij}}{p}$.
- Conversely, we could obtain MLE for $\boldsymbol{\Sigma}$ based on that for factors of $\boldsymbol{\rho}$

$$\hat{\ell}_{ij} = \hat{\ell}_{\mathbf{z},ij}\sqrt{\hat{\sigma}_{ii}}, \quad \hat{\psi}_i = \hat{\psi}_{\mathbf{z},i}\hat{\sigma}_{ii}$$

Example 9.5 (Factor analysis of stock-price data using the maximum likelihood method) The stock-price data of Examples 8.5 and 9.4 were reanalyzed assuming an $m = 2$ factor model and using the *maximum likelihood method*. The estimated factor loadings, communalities, specific variances, and proportion of total (standardized) sample variance explained by each factor are in Table 9.3.[3] The corresponding figures for the $m = 2$ factor solution obtained by the *principal component method* (see Example 9.4) are also provided. The communalities corresponding to the maximum likelihood factoring of $\mathbf{R}$ are of the form [see (9-31)] $\hat{h}_i^2 = \hat{\ell}_{i1}^2 + \hat{\ell}_{i2}^2$. So, for example,

$$\hat{h}_1^2 = (.115)^2 + (.765)^2 = .58$$

A Large Sample Test for the Number of Common Factors

- Under the normality assumption, we can test the adequacy of the m common factor model, which is

$$H_0 : \mathbf{\Sigma}_{p\times p} = \mathbf{L}_{p\times m}\mathbf{L}^T_{m\times p} + \mathbf{\Psi}_{p\times p}$$

versus H_1 : $\mathbf{\Sigma}$ any other positive definite matrix.

- Note, the MLE for general $\mathbf{\Sigma}$ is $\hat{\mathbf{\Sigma}} = \mathbf{S}_n$ with maximal likelihood proportional to $|\mathbf{S}_n|^{-\frac{n}{2}} e^{-np/2}$.
- Under H_0, the MLE of $\mathbf{\Sigma}$ is $\hat{\mathbf{\Sigma}} = \hat{\mathbf{L}}\hat{\mathbf{L}}^T + \hat{\mathbf{\Psi}}$, with the maximal likelihood proportional to

$$|\hat{\mathbf{\Sigma}}|^{-\frac{n}{2}} \exp\left\{-\frac{n}{2}\mathrm{tr}\left[\hat{\mathbf{\Sigma}}^{-1}\mathbf{S}_n\right]\right\} = |\hat{\mathbf{L}}\hat{\mathbf{L}}^T + \hat{\mathbf{\Psi}}|^{-\frac{n}{2}} \exp\left\{-\frac{n}{2}\mathrm{tr}\left[(\hat{\mathbf{L}}\hat{\mathbf{L}}^T + \hat{\mathbf{\Psi}})^{-1}\mathbf{S}_n\right]\right\}$$

A Large Sample Test for the Number of Common Factors

- Now consider the likelihood ratio statistic for testing H_0

$$-2\log\Lambda = -2\log\left[\frac{\max_{\boldsymbol{\Sigma}=\mathbf{L}\mathbf{L}^T+\boldsymbol{\Psi}} L(\hat{\boldsymbol{\mu}},\Sigma)}{\max_{\boldsymbol{\Sigma}} L(\hat{\boldsymbol{\mu}},\Sigma)}\right] = -2\log\left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\mathbf{S}_n|}\right)^{-\frac{n}{2}} + n[\mathrm{tr}(\hat{\boldsymbol{\Sigma}}^{-1}) - p]$$

- Under H_0, the MLE of $\boldsymbol{\Sigma}$ is $\hat{\boldsymbol{\Sigma}} = \hat{\mathbf{L}}\hat{\mathbf{L}}^T + \hat{\boldsymbol{\Psi}}$, which yields $\mathrm{tr}(\hat{\boldsymbol{\Sigma}}^{-1}) - p = 0$.
- Therefore we have

$$-2\log\Lambda = n\log\left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\mathbf{S}_n|}\right) \dot{\sim} \chi^2(df), \quad df = [(p-m)^2 - p - m]/2.$$

- We reject H_0 at the α-level of significance if both n and $n-p$ are large and

$$(n - 1 - (2p+4m+5)/6)\log\frac{|\hat{\mathbf{L}}\hat{\mathbf{L}}^T + \hat{\boldsymbol{\Psi}}|}{|\mathbf{S}_n|} > \chi^2_{1-\alpha}(df).$$

Example 9.7 (Testing for two common factors) The two-factor maximum likelihood analysis of the stock-price data was presented in Example 9.5. The residual matrix there suggests that a two-factor solution may be adequate. Test the hypothesis $H_0: \mathbf{\Sigma} = \mathbf{LL}' + \mathbf{\Psi}$, with $m = 2$, at level $\alpha = .05$.

The test statistic in (9-39) is based on the ratio of generalized variances

$$\frac{|\hat{\mathbf{\Sigma}}|}{|\mathbf{S}_n|} = \frac{|\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\mathbf{\Psi}}|}{|\mathbf{S}_n|}$$

Table of Contents

Factor Rotation

- Previously, we knew that all factor loadings reproduce the same covariance up to an orthogonal matrix, which defines a rotation transformation.
- Such transformation acting on the factors is called *factor rotation*.
- Suppose the $\hat{\mathbf{L}}_{p\times m}$ is the factor loading matrix obtained from data $\mathbf{X}_{n\times p}$.
- $\hat{\mathbf{L}}^* = \hat{\mathbf{L}}\mathbf{T}$ for any orthogonal matrix $\mathbf{T}_{m\times m}$ is a "rotated" loading matrix that yields the same estimated covariance

$$\hat{\mathbf{L}}\hat{\mathbf{L}}^T + \hat{\mathbf{\Psi}} = \hat{\mathbf{L}}^*\hat{\mathbf{L}}^{*T} + \hat{\mathbf{\Psi}}$$

- The residual matrix $\mathbf{S}_n - \hat{\mathbf{L}}\hat{\mathbf{L}}^T - \hat{\mathbf{\Psi}} = \mathbf{S}_n - \hat{\mathbf{L}}^*\hat{\mathbf{L}}^{*T} - \hat{\mathbf{\Psi}}$ remains unchanged.
- The estimated communalities $\hat{h}_i$ and the estimated specific variances $\hat{\psi}_i$ are unaltered.

Factor Rotation

- We often want some pattern of the factors after rotation, e.g. each variable loading highly on a single factor and light-weightedly on others.
- Such transformation acting on the factors is called *factor rotation*.
- When $m = 2$, we consider the rotation parametrized by a rotation angle ϕ.

$$\mathbf{T}_{p\times 2} = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix}$$

- The desired simple structure can be determined graphically.
- For $m > 2$, we consider rotation with an analytic measure of simple structure.
- Kaiser suggests an analytic measure of simple structure known as *varimax* criterion.
- Oblique (nonorthogonal) rotations are also considered to express each variable in terms of a minimum number of factors.

Factor Rotation

Example 9.8 (A first look at factor rotation) Lawley and Maxwell [10] present the sample correlation matrix of examination scores in $p = 6$ subject areas for $n = 220$ male students. The correlation matrix is

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

and a maximum likelihood solution for $m = 2$ common factors yields the estimates in Table 9.5.

Table 9.5

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

Factor Rotation

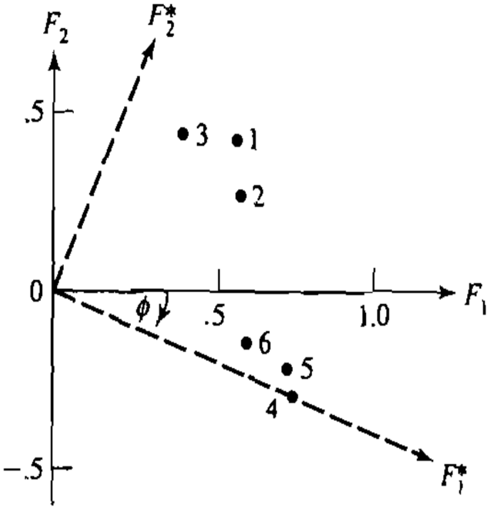

Figure 9.1 Factor rotation for test scores.

Table of Contents

Factor Scores

- In factor analysis, interest usually lies in the parameter estimation.
- However, the estimated values of the common factors, called *factor scores*, are also required.
- These quantities are often used for diagnostic purposes, and inputs to a subsequent analysis.
- Factor scores are estimates $\hat{f}_i$ of the unobserved random factor vectors $\mathbf{F}_i$, for $i = 1, \cdots, n$.
- When estimating factor scores, we
  1. treat the estimated factor loadings $\hat{\ell}_{ij}$ and the specific variances $\hat{\phi}_i$ as they were true values.
  2. (linearly) transform data (centering, standardization) and use rotated loadings to compute factor scores.

The Weighted Least Square Method

- Suppose the mean vector $\boldsymbol{\mu}$, the factor loadings $\mathbf{L}$, and the specific variance $\boldsymbol{\Psi}$ are known in the factor model

$$\mathbf{X}_{p\times 1} - \boldsymbol{\mu} = \mathbf{L}_{p\times m}\mathbf{F}_{m\times 1} + \boldsymbol{\epsilon}_{p\times 1}$$

- By assumption we have $\mathrm{Var}(\epsilon_i) = \psi_i$. Bartlett suggested the weighted least square (WLS) to estimate the common factor scores and minimized

$$\sum_{i=1}^{p} \frac{\epsilon_i^2}{\psi_i} = \boldsymbol{\epsilon}^T \boldsymbol{\Psi}^{-1} \boldsymbol{\epsilon} = (\mathbf{x} - \boldsymbol{\mu} - \mathbf{L}\mathbf{f})^T \boldsymbol{\Psi}^{-1} (\mathbf{x} - \boldsymbol{\mu} - \mathbf{L}\mathbf{f})$$

- And we get the WLSE for $\mathbf{f}$

$$\hat{\mathbf{f}} = (\mathbf{L}^T \boldsymbol{\Psi}^{-1} \mathbf{L})^{-1} \mathbf{L}^T \boldsymbol{\Psi}^{-1} (\mathbf{x} - \boldsymbol{\mu})$$

The Weighted Least Square Method

- With estimates $\hat{\mathbf{L}}$, $\hat{\boldsymbol{\Psi}}$ and $\hat{\mu} = \bar{\mathbf{x}}$, we have

$$\hat{\mathbf{f}}_i^{WLS} = (\hat{\mathbf{L}}^T \hat{\boldsymbol{\Psi}}^{-1} \hat{\mathbf{L}})^{-1} \hat{\mathbf{L}}^T \hat{\boldsymbol{\Psi}}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}})$$

- When $\hat{\mathbf{L}}$ and $\hat{\boldsymbol{\Psi}}$ are obtained by MLE, we need to assume $\hat{\mathbf{L}}^T \hat{\boldsymbol{\Psi}}^{-1} \hat{\mathbf{L}} = \hat{\Delta}$ is a diagonal matrix. Then

$$\hat{\mathbf{f}}_i^{WLS} = \hat{\Delta}^{-1} \hat{\mathbf{L}}^T \hat{\boldsymbol{\Psi}}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}), \quad i = 1, \cdots, n$$

- If the standardized variables $\mathbf{z}_i = \mathbf{V}^{-\frac{1}{2}} (\mathbf{x}_i - \bar{\mathbf{x}})$ are used, then

$$\hat{\mathbf{f}}_i^{WLS} = \hat{\Delta}_{\mathbf{z}}^{-1} \hat{\mathbf{L}}_{\mathbf{z}}^T \hat{\boldsymbol{\Psi}}_{\mathbf{z}}^{-1} \mathbf{z}_i, \quad i = 1, \cdots, n$$

  where $\hat{\rho} = \hat{\mathbf{L}}_{\mathbf{z}} \hat{\mathbf{L}}_{\mathbf{z}}^T + \hat{\boldsymbol{\Psi}}_{\mathbf{z}}$.
- If a rotated loading matrix $\hat{\mathbf{L}}^* = \hat{\mathbf{L}}\mathbf{T}$ is used, then the corresponding factor scores can be obtained as $\hat{\mathbf{f}}_i^* = \mathbf{T}^T \hat{\mathbf{f}}_i$ for $i = 1, \cdots, n$.

The Regression Method

- Now consider the original factor model with $\mathbf{F} \sim N_m(\mathbf{0}, \mathbf{I}_m)$ and $\epsilon \sim N_p(\mathbf{0}, \mathbf{\Psi})$

$$\mathbf{X} = \boldsymbol{\mu} + \mathbf{LF} + \epsilon \sim N_p(\boldsymbol{\mu}, \mathbf{\Sigma}), \quad \mathbf{\Sigma} = \mathbf{LL}^T + \mathbf{\Psi}$$

- When $\mathbf{F}$ and ϵ are assumed jointly normal, then we have

$$\begin{bmatrix} \mathbf{X} - \boldsymbol{\mu} \\ \mathbf{F} \end{bmatrix} \sim N_{m+p}(\mathbf{0}, \mathbf{\Sigma}^*), \quad \mathbf{\Sigma}^* = \begin{bmatrix} \mathbf{\Sigma} = \mathbf{LL}^T + \mathbf{\Psi} & \mathbf{L} \\ \mathbf{L}^T & \mathbf{I}_m \end{bmatrix}$$

- Then we get the conditional normal for $\mathbf{F}|\mathbf{x}$ with

$$\mathrm{E}(\mathbf{F}|\mathbf{x}) = \mathbf{L}^T\mathbf{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = \mathbf{L}^T(\mathbf{LL}^T + \mathbf{\Psi})^{-1}(\mathbf{x} - \boldsymbol{\mu})$$
$$\mathrm{Cov}(\mathbf{F}|\mathbf{x}) = \mathbf{I} - \mathbf{L}^T\mathbf{\Sigma}^{-1}\mathbf{L} = \mathbf{I} - \mathbf{L}^T(\mathbf{LL}^T + \mathbf{\Psi})^{-1}\mathbf{L} = (\mathbf{I} + \mathbf{L}^T\mathbf{\Psi}^{-1}\mathbf{L})^{-1}$$

The Regression Method

- With estimates $\hat{\mathbf{L}}$, $\hat{\mathbf{\Psi}}$ and $\hat{\mu} = \bar{\mathbf{x}}$, we have

$$\hat{\mathbf{f}}_i^R = \hat{\mathbf{L}}^T \hat{\mathbf{\Sigma}}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}) = \hat{\mathbf{L}}^T(\hat{\mathbf{L}}\hat{\mathbf{L}}^T + \hat{\mathbf{\Psi}})^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}) = (\mathbf{I} + \hat{\mathbf{L}}^T \hat{\mathbf{\Psi}}^{-1}\hat{\mathbf{L}})^{-1}\hat{\mathbf{L}}^T \hat{\mathbf{\Psi}}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})$$

- A quick comparison yields the connection between two solutions

$$(\hat{\mathbf{L}}^T \hat{\mathbf{\Psi}}^{-1}\hat{\mathbf{L}})\hat{\mathbf{f}}_i^{WLS} = (\mathbf{I} + \hat{\mathbf{L}}^T \hat{\mathbf{\Psi}}^{-1}\hat{\mathbf{L}})\hat{\mathbf{f}}_i^R$$

or equivalently $\hat{\mathbf{f}}_i^{WLS} = (\mathbf{I} + \hat{\Delta}^{-1})\hat{\mathbf{f}}_i^R$.

- Practitioners tend to use $\hat{\Sigma} = \mathbf{S}$. Then we have

$$\hat{\mathbf{f}}_i^R = \hat{\mathbf{L}}^T \mathbf{S}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}), \quad i = 1, \cdots, n$$

- If the standardized variables $\mathbf{z}_i = \mathbf{V}^{-\frac{1}{2}}(\mathbf{x}_i - \bar{\mathbf{x}})$ are used, then

$$\hat{\mathbf{f}}_i^R = \hat{\mathbf{L}}_{\mathbf{z}}^T \boldsymbol{P}^{-1}\mathbf{z}_i, \quad i = 1, \cdots, n$$

Example 9.12 (Computing factor scores) We shall illustrate the computation of factor scores by the least squares and regression methods using the stock-price data discussed in Example 9.10. A maximum likelihood solution from $\mathbf{R}$ gave the estimated rotated loadings and specific variances

$$\hat{\mathbf{L}}_{\mathbf{z}}^{*} = \begin{bmatrix} .763 & .024 \\ .821 & .227 \\ .669 & .104 \\ .118 & .993 \\ .113 & .675 \end{bmatrix} \quad \text{and} \quad \hat{\boldsymbol{\Psi}}_{\mathbf{z}} = \begin{bmatrix} .42 & 0 & 0 & 0 & 0 \\ 0 & .27 & 0 & 0 & 0 \\ 0 & 0 & .54 & 0 & 0 \\ 0 & 0 & 0 & .00 & 0 \\ 0 & 0 & 0 & 0 & .53 \end{bmatrix}$$

The vector of standardized observations,

$$\mathbf{z}' = [.50, -1.40, -.20, -.70, 1.40]$$

yields the following scores on factors 1 and 2: