多元统计分析

第6讲 因子分析

Johnson & Wichern 9.1-9.6, Supplement 9A

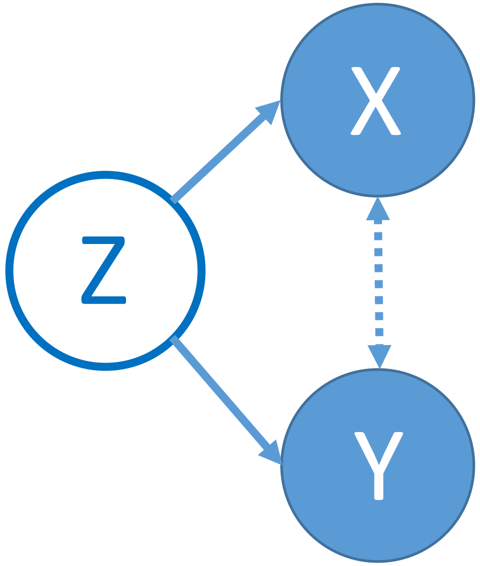
统计学研究中心 侯琳
houl@tsinghua.edu.cn

2016-2017春季学期

Overview of Factor Analysis

- Early development in psychometrics by Karl Pearson, Charles Spearman, etc
- To describe the covariance structure among many variables with a few unobservable or latent variables called factors
  - Reduction: reduce high dimension data to a few variables
  - Interpretation: explain the covariance of observed variables with latent factors

$$\underset{(p\times p)}{\boldsymbol{\Sigma}} = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{bmatrix}$$



Orthogonal Factor Model (I)

- The observable random vector **X**, with *p* components, with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$
- X is linearly dependent upon a few common factors and specific factors, with

loading

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

or

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

Orthogonal Factor Model (II)

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

$$E(\mathbf{F}) = 0, \quad \text{Cov}(\mathbf{F}) = E(\mathbf{F}\mathbf{F}') = \underset{(m\times m)}{\mathbf{I}}$$

$$E(\boldsymbol{\varepsilon}) = 0, \quad \text{Cov}(\boldsymbol{\varepsilon}) = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \underset{(p\times p)}{\boldsymbol{\Psi}}$$

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

$$\boldsymbol{\Psi} = \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_p \end{bmatrix}$$

Covariance Structure Implied

$$(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})' = (\mathbf{LF}+\boldsymbol{\varepsilon})((\mathbf{LF}+\boldsymbol{\varepsilon})'$$
$$= (\mathbf{LF})(\mathbf{F}'\mathbf{L}') + \boldsymbol{\varepsilon}(\mathbf{F}'\mathbf{L}') + (\mathbf{LF})\boldsymbol{\varepsilon}' + \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'$$

$$\boldsymbol{\Sigma} = \mathrm{Cov}(\mathbf{X}) = E(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})'$$
$$= \mathbf{L}E(\mathbf{FF}')\mathbf{L}' + E(\boldsymbol{\varepsilon}\mathbf{F}')\mathbf{L}' + \mathbf{L}E(\mathbf{F}\boldsymbol{\varepsilon}') + E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')$$
$$= \mathbf{LL}' + \boldsymbol{\Psi}$$

$$\sigma_{ii} = l_{i1}^2 + l_{i2}^2 + \cdots + l_{im}^2 + \psi_i$$
$$= h_i^2 + \psi_i$$

Communality + Specific variance

$$\sigma_{ik} = \mathrm{Cov}(X_i, X_k) = \mathbf{l}_i'\mathbf{l}_k$$
$$= l_{i1}l_{k1} + \cdots + l_{im}l_{km}$$

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

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

Discussions (I)

Reduction

- How many parameters are there in a covariance matrix?
- How many parameters are there in the orthogonal factor model?
- What is the maximum number of common factors?

Note: Not all covariance matrix can be factored as $\mathbf{LL}'+\boldsymbol{\Psi}$, where the number of factors $m << p$

See Example 9.2 in textbook

Discussions (II)

Uniqueness

- Consider orthogonal matrix **T**

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

Check model assumptions

$$=\mathbf{L}(\mathbf{TT}')\mathbf{F}+\boldsymbol{\varepsilon}$$

L is not unique!

$$=(\mathbf{LT})(\mathbf{T}'\mathbf{F})+\boldsymbol{\varepsilon}$$

$$\mathbf{L}^*=\mathbf{LT},\quad \mathbf{F}^*=\mathbf{T}'\mathbf{F}$$

The communalities are not affected by choices of **T**

$$\mathbf{X}-\boldsymbol{\mu}=\mathbf{L}^*\mathbf{F}^*+\boldsymbol{\varepsilon}$$

- Factor rotation

Textbook 9.1, 9.2

Methods of Estimation

- Suppose $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ represent *n* independent drawings from some p-dimensional population, with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.
- Sample covariance matrix **S,** sample correlation matrix **R**
- Objective: find $\hat{\mathbf{L}}$ and $\hat{\boldsymbol{\Psi}}$, with $\mathbf{S} \approx \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}}$

$$\boldsymbol{\Sigma} = \lambda_1 \mathbf{e}_1 \mathbf{e}_1' + \lambda_2 \mathbf{e}_2 \mathbf{e}_2' + \cdots + \lambda_p \mathbf{e}_p \mathbf{e}_p'$$

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

$$\underset{(p\times p)}{\boldsymbol{\Sigma}} = \underset{(p\times p)}{\mathbf{L}}\;\underset{(p\times p)}{\mathbf{L}'} + \underset{(p\times p)}{\mathbf{0}} = \mathbf{L}\mathbf{L}'$$

The spectral decomposition is not useful!
common factors = # variables

Estimation: Principal Component Approach

- When the last *p-m* eigenvalues are small, neglect the contribution of the corresponding eigenvalue-eigenvector pairs

$$\Sigma \doteq \left[\sqrt{\lambda_1}\,\mathbf{e}_1 \;\vdots\; \sqrt{\lambda_2}\,\mathbf{e}_2 \;\vdots\; \cdots \;\vdots\; \sqrt{\lambda_m}\,\mathbf{e}_m\right] \begin{bmatrix} \sqrt{\lambda_1}\,\mathbf{e}_1' \\ \cdots \\ \sqrt{\lambda_2}\,\mathbf{e}_2' \\ \cdots \\ \vdots \\ \cdots \\ \sqrt{\lambda_m}\,\mathbf{e}_m' \end{bmatrix} = \underset{(p\times m)}{\mathbf{L}}\;\underset{(m\times p)}{\mathbf{L}'}$$

What is $\tilde{\Psi}$?

What is communality $\tilde{h}_i^2$?

Not diagonal

The specific variances may be taken to be the diagonal elements of $\Sigma - \tilde{\mathbf{L}}\tilde{\mathbf{L}}'$

Note: the estimated loadings for a given factor do not change as the number of factors is increased.

Selection of *m*

Matrix Approximation

- Similar to PCA
- Consider the residual matrix $\mathbf{S}-(\tilde{\mathbf{L}}\tilde{\mathbf{L}}'+\tilde{\mathbf{\Psi}})$
- It can be shown that

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

- Consequently, a small value for the sum of the squares of the neglected eigenvalues implies a small value for the sum of the squared errors of approximation.

Selection of m: another perspective

- What is the total variance in $\mathbf{X}$?
- What is the contribution of common factor i to the total variance?
- Proportion of variance explained by the common factors

Example 9.3: Factor Analysis of CP Data

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

What is the maximum *m*?

Table 9.1

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		
2. Good buy for money	.78	−.53		
3. Flavor	.65	.75		
4. Suitable for snack	.94	−.10		
5. Provides lots of energy	.80	−.54		
Eigenvalues				
Cumulative proportion of total (standardized) sample variance				

Example 9.3: Factor Analysis of CP Data

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

What is the maximum *m*?

Table 9.1

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		

Estimation: Principal Factor Solution

- Intuitive idea: the common factors should account for the off-diagonal elements, as well as the communality portions of the diagonal elements

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

$$\boldsymbol{\Sigma} = \mathbf{LL}' + \boldsymbol{\Psi}$$

Initial $\tilde{\boldsymbol{\Psi}}$

Find $\tilde{\mathbf{L}}$, the largest m eigenvectors of the

eigen decomposition of $\mathbf{S} - \tilde{\boldsymbol{\Psi}}$

$$\tilde{\boldsymbol{\Psi}} = diag(\mathbf{S} - \tilde{\mathbf{L}}\tilde{\mathbf{L}}')$$

Discussions

- Choice of initial estimates of specific variances
- Some of the eigenvalues of $\mathbf{S} - \tilde{\mathbf{\Psi}}$ may be negative
- Communality may exceed total variance, Heywood case
- Reasonable suggestion of how to initial $\tilde{\mathbf{\Psi}}$

$$h_i^{*2} = 1 - \psi_i^* = 1 - \frac{1}{r^{ii}}$$

r^{ii} is the i-th diagonal element of $\mathbf{R}^{-1}$

Estimation: Maximum Likelihood Method

- Assumption: the common factors and the specific factors are jointly normally distributed

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

$$\mathbf{F} \sim N_m(0, \mathbf{I})$$

$$\boldsymbol{\varepsilon} \sim N_p(0, \boldsymbol{\Psi})$$

$\Rightarrow$

$$\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

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

subject to $\mathbf{L}'\boldsymbol{\Psi}^{-1}\mathbf{L} = \boldsymbol{\Delta}$

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

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

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

The model depends on $\mathbf{L}$ and $\boldsymbol{\Psi}$ through $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}$
It is not well defined because of multiplicity of choices of $\mathbf{L}$

Impose computationally convenient uniqueness condition:

$\mathbf{L}'\boldsymbol{\Psi}^{-1}\mathbf{L} = \boldsymbol{\Delta}$, $\boldsymbol{\Delta}$ is a diagonal matrix

R function, factanal()

- Result 9.1 Let $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ be a random sample from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}$ is the covariance matrix for the *m* common factor model. The maximum likelihood estimators $\hat{\mathbf{L}}, \hat{\boldsymbol{\Psi}}$, and $\hat{\boldsymbol{\mu}}$ subject to $\hat{\mathbf{L}}\hat{\boldsymbol{\Psi}}^{-1}\hat{\mathbf{L}}'$ being diagonal.

  Then, the MLE of the communalities are

$$\hat{h}_i^2 = \hat{l}_{i1}^2 + \hat{l}_{i2}^2 + \cdots + \hat{l}_{im}^2, \text{ for i} = 1,2,\cdots,\text{p}$$

so

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

- If the variables are standardized so that $Z = \mathbf{V}^{-\frac{1}{2}}(\mathbf{X} - \boldsymbol{\mu})$
- What is the covariance matrix $\boldsymbol{\rho}$?

$$\boldsymbol{\rho} = \mathbf{V}^{-1/2}\boldsymbol{\Sigma}\mathbf{V}^{-1/2} = (\mathbf{V}^{-1/2}\mathbf{L})(\mathbf{V}^{-1/2}\mathbf{L})' + \mathbf{V}^{-1/2}\boldsymbol{\Psi}\mathbf{V}^{-1/2}$$

- Thus, we have a factorization of $\boldsymbol{\rho}$:

$$\mathbf{L}_z = \mathbf{V}^{-1/2}\mathbf{L},\ \boldsymbol{\Psi}_z = \mathbf{V}^{-1/2}\boldsymbol{\Psi}\mathbf{V}^{-1/2}$$

- The MLE of $\boldsymbol{\rho}$ is

$$\hat{\boldsymbol{\rho}} = (\hat{\mathbf{V}}^{-1/2}\hat{\mathbf{L}})(\hat{\mathbf{V}}^{-1/2}\hat{\mathbf{L}})' + \hat{\mathbf{V}}^{-1/2}\hat{\boldsymbol{\Psi}}\hat{\mathbf{V}}^{-1/2}$$

$$= \hat{\mathbf{L}}_z\hat{\mathbf{L}}_z' + \hat{\boldsymbol{\Psi}}_z$$

Question:

Why?

What is $\hat{\mathbf{V}}^{-1/2}$?

Does PC approach have similar property?

Note:

- The MLE method could produce very different results when m→ m+1
- The MLE method can also experience difficulties with Heywood cases

factanal

Example 9.5 Factor Analysis of stock-price data

Table 9.3

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

Discussion:
- Are the columns orthogonal?
- Estimated value
- % of total variance explained?

Homework:
princomp()
factanal()

A Large Sample Test for the Number of Common Factors

- Normality Assumption: the common factors and the specific factors are jointly normally distributed

$$H_0: \underset{(p\times p)}{\mathbf{\Sigma}} = \underset{(p\times m)}{\mathbf{L}} \underset{(m\times p)}{\mathbf{L}'} + \underset{(p\times p)}{\mathbf{\Psi}}, \text{subject to } \mathbf{L}'\mathbf{\Psi}\mathbf{L} = \mathbf{\Delta}$$

$$H_1: \ \mathbf{\Sigma} \text{ any other positive definte matrix}$$

- Likelihood Ratio test $-2\ln\Lambda = -2\ln\left[\dfrac{\text{maximum likelihood under } H_0}{\text{maximized likelihood}}\right]$

- Under alternative, what is the MLE estimator?

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}},\ \hat{\boldsymbol{\Sigma}} = \frac{n-1}{n}\mathbf{S}, \text{or } \mathbf{S}_n$$

$$\text{the maximized likelihood} \propto |\mathbf{S}_n|^{-n/2} e^{-np/2}$$

$$-2\ln\Lambda = -2\ln\left[\frac{\text{maximum likelihood under } H_0}{\text{maximized likelihood}}\right] = n\ln\left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\mathbf{S}_n|}\right)$$

Degree of Freedom :

$$\nu - \nu_0 = \frac{1}{2}p(p+1) - p(m+1) + \frac{1}{2}m(m-1)$$

- Under null:

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$$

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

$$\text{the maximized likelihood} \propto |\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}}|^{-n/2} \exp\{-\frac{1}{2} n \operatorname{tr}[(\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}})^{-1}\mathbf{S}_n]\}$$

Factor Rotation

- Factor loading is not unique
- Initial loading + orthogonal transformation

$$\hat{\mathbf{L}}^* = \hat{\mathbf{L}}\mathbf{T}, \text{ with } \mathbf{T}\mathbf{T}' = \mathbf{T}'\mathbf{T} = \mathbf{I}$$

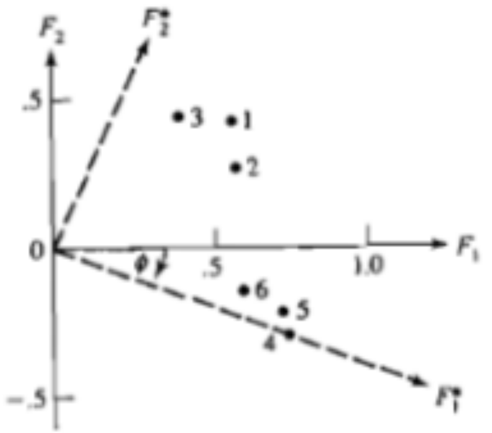
- Question
  - Is the covariance/correlation matrix changed after rotation? What about residual matrix, estimated specific variances, communalities?
  - Why rotation? (Interpretation, PC approach)
  - Criteria? What is a desirable result?

What is "simpler" structure?

- Ideally, we should like to see a pattern of loadings such that each variable loads highly on a single factor and has small to moderate loadings on the remaining factors.

Table 9.7

Variable	Estimated factor loadings F_1	Estimated factor loadings F_2	Rotated estimated factor loadings F_1^*	Rotated estimated factor loadings F_2^*	Communalities $\tilde{h}_i^2$
1. Taste	.56	.82	.02	.99	.98
2. Good buy for money	.78	−.52	.94	−.01	.88
3. Flavor	.65	.75	.13	.98	.98
4. Suitable for snack	.94	−.10	.84	.43	.89
5. Provides lots of energy	.80	−.54	.97	−.02	.93
Cumulative proportion of total (standardized) sample variance explained	.571	.932	.507	.932	



Varimax Criterion

$$\tilde{l}_{ij}^* = \hat{l}_{ij}^* / \hat{h}_i$$

Varimax procedure selects the orthogonal transformation **T** that maximizes

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

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

Factor Score

- Recall: what are the scores in principal component analysis?
- Factor score v.s. PC score
- Weighted least squares method

Question:
- What is the factor scores if MLE method is used?
- What if the correlation matrix is factored?

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

$$\sum_{i=1}^{p}\frac{\varepsilon_i^2}{\psi_i} = \boldsymbol{\varepsilon}^{-1}\boldsymbol{\Psi}^{-1}\boldsymbol{\varepsilon} = (\mathbf{x}-\boldsymbol{\mu}-\mathbf{Lf})'\boldsymbol{\Psi}^{-1}(\mathbf{x}-\boldsymbol{\mu}-\mathbf{Lf}) \qquad (1)$$

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

Take $\hat{\mathbf{L}}, \hat{\boldsymbol{\Psi}}$, and $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$ as the true value to obtain the factor score

Factor Score – Regression Method

- Recall: multivariate normal distribution, conditional distribution

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

$$\mathbf{X} \sim N_p(\boldsymbol{\mu}, \mathbf{LL}' + \boldsymbol{\Psi})$$

- Q: the joint distribution of $(\mathbf{X} - \boldsymbol{\mu}, \mathbf{F})$?
- Q: the conditional distribution $\mathbf{F} \mid \mathbf{X}$?

$$\begin{pmatrix} \mathbf{X}-\boldsymbol{\mu} \\ \mathbf{F} \end{pmatrix} = \begin{pmatrix} \mathbf{L} \\ \mathbf{I}_m \end{pmatrix} \mathbf{F} + \begin{pmatrix} \boldsymbol{\varepsilon} \\ \mathbf{0} \end{pmatrix}$$

$$\mathbf{F} \sim N_m(\mathbf{0}, \mathbf{I})$$

$$\begin{pmatrix} \mathbf{X}-\boldsymbol{\mu} \\ \mathbf{F} \end{pmatrix} \sim N_{p+m}\left(\mathbf{0}, \begin{bmatrix} \mathbf{LL}'+\boldsymbol{\Psi} & \mathbf{L} \\ \mathbf{L}' & \mathbf{I}_m \end{bmatrix} \right)$$

Relationship to mean structure of response in regression analysis

$$\text{mean} = \mathrm{E}(\mathbf{F} \mid \mathbf{x}) = \mathbf{L}'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = \mathbf{L}'(\mathbf{LL}'+\boldsymbol{\Psi})^{-1}(\mathbf{x}-\boldsymbol{\mu})$$

$$\text{covariance} = \mathrm{Cov}(\mathbf{F} \mid \mathbf{x}) = \mathbf{I}_m - \mathbf{L}'\boldsymbol{\Sigma}^{-1}\mathbf{L} = \mathbf{I}_m - \mathbf{L}'(\mathbf{LL}'+\boldsymbol{\Psi})^{-1}\mathbf{L}$$

The j-th factor score vector is given by

$$\hat{\mathrm{f}}_j = \hat{\mathbf{L}}'\hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}}) = \hat{\mathbf{L}}'(\hat{\mathbf{L}}\hat{\mathbf{L}}'+\hat{\boldsymbol{\Psi}})^{-1}(\mathbf{x}_j - \bar{\mathbf{x}})$$

Discussion (I)

(1)Regression method and WLS method

$$\hat{\mathbf{f}}^{R}_{j} = \hat{\mathbf{L}}'\hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}}) = \hat{\mathbf{L}}'(\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}})^{-1}(\mathbf{x}_j - \bar{\mathbf{x}})$$

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

$$\hat{\mathbf{L}}'(\hat{\mathbf{L}}'\hat{\mathbf{L}} + \hat{\boldsymbol{\Psi}})^{-1} = (\mathbf{I} + \hat{\mathbf{L}}'\hat{\boldsymbol{\Psi}}^{-1}\hat{\mathbf{L}})^{-1}\hat{\mathbf{L}}'\hat{\boldsymbol{\Psi}}^{-1}$$ (HW, see excercise 9.6)

$$\hat{\mathbf{f}}_j^{WLS} = (\mathbf{I} + (\hat{\mathbf{L}}'\hat{\boldsymbol{\Psi}}^{-1}\hat{\mathbf{L}})^{-1})\hat{\mathbf{f}}_j^{R}$$

(2) $\mathbf{S}$ is often used for $\hat{\boldsymbol{\Sigma}}$ rather than $\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}}$

$$\hat{\mathbf{f}}_j^{R} = \hat{\mathbf{L}}'\mathbf{S}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}})$$

(3) $\hat{\mathbf{f}}_j^{R} = \hat{\mathbf{L}}_z'\mathbf{R}^{-1}\mathbf{z}_j$

Discussion (II)

- Factor analysis versus Principal Component Analysis

$$\hat{\mathbf{f}}_j^R = \hat{\mathbf{L}}'\mathbf{S}^{-1}(\mathbf{x_j} - \bar{\mathbf{x}})$$

- Consider PC approach to estimate the factor loadings

Since $\mathbf{S} = \mathbf{P\Lambda\Lambda}', \mathbf{L} = \mathbf{P\Lambda}^{1/2}$

$$\hat{\mathbf{f}}_j^R = \hat{\mathbf{L}}'\mathbf{S}^{-1}(\mathbf{x_j} - \bar{\mathbf{x}}) = (\mathbf{P\Lambda}^{1/2})'(\mathbf{P\Lambda}^{-1}\mathbf{P}')(\mathbf{x_j} - \bar{\mathbf{x}})$$

$$= \mathbf{\Lambda}^{-1/2}\boxed{\mathbf{P}'(\mathbf{x_j} - \bar{\mathbf{x}})}$$

PC score

Discussion (III)

- Factor rotation $\hat{\mathbf{L}}^* = \hat{\mathbf{L}}\mathbf{T}, \hat{\mathbf{f}}_j^* = \mathbf{T}'\hat{\mathbf{f}}_j$

- Strategy for factor analysis
  1. Perform a principal component factor analysis
  2. Perform a maximum likelihood factor analysis
  3. Compare the solutions obtained from the two factor analyses
  4. Repeat the steps 1-3 for the other number of common factors *m*
  5. For large data sets, split them in half and perform FA on each part

Factor Analysis

9.2 Orthogonal Factor Model
- Common factor
- Specific factor
- Factor loading
- Communality

9.3 Methods of Estimation
- PC method
- Principal factor solution
- Maximum likelihood method

9.4: Factor Rotation
- Varimax criteria

9.5: Factor Scores
- Weighted least squares method
- Regression method

Discussions:
PCA and Factor Analysis