Factor Analysis and Inference for Structured Covariance Matrices

Shyh-Kang Jeng
Department of Electrical Engineering/
Graduate Institute of Communication/
Graduate Institute of Networking and Multimedia

1

Outline

- Introduction
- The Orthogonal Factor Model
- Methods of Estimation
- Factor Rotation
- Factor Scores
- Perspectives and Strategy for Factor Analysis

Outline

- Introduction
- The Orthogonal Factor Model
- Methods of Estimation
- Factor Rotation
- Factor Scores
- Perspectives and Strategy for Factor Analysis

Questions

- What is the essential purpose of factor analysis?
- What is a factor?
- What is the difference between the principal component analysis and the factor analysis?

History

- Early 20th-century attempt to define and measure intelligence
- Developed primarily by scientists interested in psychometrics
- Advent of computers generated a renewed interest
- Each application must be examined on its own merits

5

Essence of Factor Analysis

- Describe the covariance among many variables in terms of a few underlying, but unobservable, random *factors*.
- A group of variables highly correlated among themselves, but having relatively small correlations with variables in different groups represent a single underlying *factor*

6

Example 9.8
Examination Scores

R =

	Gaelic	English	History	Arithmetic	Algebra	Geometry
	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

7

Outline

- Introduction
- The Orthogonal Factor Model
- Methods of Estimation
- Factor Rotation
- Factor Scores
- Perspectives and Strategy for Factor Analysis

Questions

- What is the system of equations for a factor analysis model?
- What is the orthogonal factor model?
- What is the communality?
- What is the specific variance?
- Is the factor model unique? Why?

Orthogonal Factor Model

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

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

ℓ_{ij} : loading of the ith variable on the jth factor

$F_1, F_2, \cdots, F_m, \varepsilon_1, \varepsilon_2, \cdots, \varepsilon_p$:

unobservable random variables

10

Orthogonal Factor Model

$\mathbf{X}$: random vector with p components

$\boldsymbol{\mu}, \boldsymbol{\Sigma}$: mean and covariance matrix of $\mathbf{X}$

$\mathbf{F}$: common factors, independent of $\mathbf{X}$

random vector with m components

$\boldsymbol{\varepsilon}$: errors or specific factors, source of variation

random vector with p components

$\mathbf{L}$: matrix of factor loading

11

Orthogonal Factor Model

$$E(\mathbf{F}) = 0, \quad \mathrm{Cov}(\mathbf{F}) = E(\mathbf{F}\mathbf{F}') = \mathbf{I}$$
$$E(\boldsymbol{\varepsilon}) = 0, \quad \mathrm{Cov}(\boldsymbol{\varepsilon}) = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \boldsymbol{\Psi}$$
$$\boldsymbol{\Psi} = \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_p \end{bmatrix}$$
$$\mathrm{Cov}(\boldsymbol{\varepsilon}, \mathbf{F}) = E(\boldsymbol{\varepsilon}\mathbf{F}') = \mathbf{0}$$

12

Orthogonal Factor Model

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

13

Orthogonal Factor Model

$$\mathrm{Cov}(\mathbf{X}) = \mathbf{LL}'+\boldsymbol{\Psi}$$
$$\mathrm{Var}(X_i) = \sigma_{ii} = \ell_{i1}^2+\ell_{i2}^2+\cdots+\ell_{im}^2+\psi_i$$
$$\text{communality} = h_i^2 = \ell_{i1}^2+\ell_{i2}^2+\cdots+\ell_{im}^2$$
$$\mathrm{Var}(X_i) = \text{communality} + \text{specific variance}$$
$$\mathrm{Cov}(X_i,X_k) = \ell_{i1}\ell_{k1}+\cdots+\ell_{im}\ell_{km}$$
$$\mathrm{Cov}(\mathbf{X},\mathbf{F}) = \mathbf{L}$$
$$\mathrm{Cov}(X_i,F_j) = \ell_{ij}$$

14

Example 9.1: Verification

$$\boldsymbol{\Sigma} = \begin{bmatrix} 19 & 30 & 2 & 12 \\ 30 & 57 & 5 & 23 \\ 2 & 5 & 38 & 47 \\ 12 & 23 & 47 & 68 \end{bmatrix}$$
$$= \begin{bmatrix} 4 & 1 \\ 7 & 2 \\ -1 & 6 \\ 1 & 8 \end{bmatrix}\begin{bmatrix} 4 & 7 & -1 & 1 \\ 1 & 2 & 6 & 8 \end{bmatrix}+\begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$
$$h_1^2 = 4^2+1^2 = 17, \quad \sigma_{11}^2 = 19 = 17+2 = h_1^2+\psi_1^2$$

15

Example 9.2: No Solution

$$X_1-\mu_1 = \ell_{11}F_1+\varepsilon_1,\ X_2-\mu_2 = \ell_{21}F_1+\varepsilon_2,\ X_3-\mu_3 = \ell_{31}F_1+\varepsilon_3$$
$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.9 & 0.7 \\ 0.9 & 1 & 0.4 \\ 0.7 & 0.4 & 1 \end{bmatrix} = \begin{bmatrix} \ell_{11} \\ \ell_{21} \\ \ell_{31} \end{bmatrix}\begin{bmatrix} \ell_{11} & \ell_{21} & \ell_{31} \end{bmatrix}+\begin{bmatrix} \psi_1 & 0 & 0 \\ 0 & \psi_2 & 0 \\ 0 & 0 & \psi_3 \end{bmatrix}$$
$$1 = \ell_{11}^2+\psi_1, \quad 0.9 = \ell_{11}\ell_{21}, \quad 0.7 = \ell_{11}\ell_{31}$$
$$1 = \ell_{21}^2+\psi_2, \quad 0.4 = \ell_{21}\ell_{31}, \quad 1 = \ell_{31}^2+\psi_3$$
$$\ell_{21} = (0.4/0.7)\ell_{11}, \quad \ell_{11}^2 = 1.575, \quad \mathrm{Cov}(F_1) = 1, \mathrm{Cov}(X_1) = 1$$
$$|\ell_{11}| = |\mathrm{Cov}(X_1,F_1)| = |\mathrm{Corr}(X_1,F_1)| \le 1, \text{contradiction}$$
$$\psi_1 = 1-\ell_{11}^2 = -0.575 = \mathrm{Var}(\varepsilon_1), \text{unsatisfactory}$$

16

Ambiguities of $\mathbf{L}$ When $m>1$

$\mathbf{T}: m \times m$ orthogonal matrix, $\mathbf{TT'} = \mathbf{I}$

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

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

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

$\boldsymbol{\Sigma} = \mathbf{LL}' + \boldsymbol{\Psi} = \mathbf{LTT'L}' + \boldsymbol{\Psi} = (\mathbf{L}^*)(\mathbf{L}^*)' + \boldsymbol{\Psi}$

communalities are unaffected by the choice of $\mathbf{T}$

orthogonal matrix $\mathbf{T}$: rotation of factors in the coordinate system for $\mathbf{X}$

17

Outline

- Introduction
- The Orthogonal Factor Model
- Methods of Estimation
- Factor Rotation
- Factor Scores
- Perspectives and Strategy for Factor Analysis

Questions

- How to use the principal component analysis to solve for the factors?
- What is the residual matrix?
- What is the upper limit for the sum of squared entries of the residual matrix?
- How to determine the number of common factors?

Questions

- How to apply the maximum likelihood method to solve for the factors?
- What is the maximum likelihood estimates of the communalities?
- What is the proportion of total sample variance due to the jth factor?

Questions

- How to conduct a large sample test for the number of common factors?

Principal Component Solution

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

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

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

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

Principal Component Solution

$(\hat{\lambda}_i, \hat{\mathbf{e}}_i)$: eigenvalue-eigenvector pairs of $\mathbf{S}$ or $\mathbf{R}$

$\hat{\lambda}_1 \ge \hat{\lambda}_2 \ge \cdots \ge \hat{\lambda}_p, \quad m < p$

matrix of estimated factor loadings:

$$\tilde{\mathbf{L}} = \{\tilde{\ell}_{ij}\} = \left[\sqrt{\hat{\lambda}_1}\hat{\mathbf{e}}_1 \quad \sqrt{\hat{\lambda}_2}\hat{\mathbf{e}}_2 \quad \cdots \quad \sqrt{\hat{\lambda}_m}\hat{\mathbf{e}}_m\right]$$

estimated specific variance:

$$\tilde{\boldsymbol{\Psi}} = diag\{\psi_1, \psi_2, \cdots, \psi_p\}, \psi_i = s_{ii} - \sum_{j=1}^{m} \tilde{\ell}_{ij}^2$$

estimated communalities: $\tilde{h}_i^2 = \tilde{\ell}_{i1}^2 + \tilde{\ell}_{i2}^2 + \cdots + \tilde{\ell}_{im}^2$

Residual Matrix

$\mathbf{S} - (\tilde{\mathbf{L}}\tilde{\mathbf{L}}' + \tilde{\boldsymbol{\Psi}})$: residual matrix

sum of squared entities of residual matrix

$\le \hat{\lambda}_{m+1}^2 + \cdots + \hat{\lambda}_p^2$

contribution to s_{ii} from the first common factor: $\tilde{\ell}_{i1}^2$

contribution to the total sample variance

$s_{11} + s_{22} + \cdots + s_{pp} = \text{tr}(\mathbf{S})$

from the first common factor:

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

Determination of Number of Common Factors

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

number of common factors retained is increased until a "suitable proportion" of the total sample variance has been explained

Another criterion : set m equal to the number of eigenvalues of $\mathbf{R}$ greater than 1 if $\mathbf{R}$ is factored

25

Example 9.3 Consumer Preference 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

26

Example 9.3 Determination of m

$\hat{\lambda}_1 = 2.85$, $\hat{\lambda}_2 = 1.81$ are the only eigenvalues greater than 1

They account for a cumulative proportion

$$\frac{\hat{\lambda}_1 + \hat{\lambda}_2}{p} = 0.93$$

of the total (standard) sample variance

$\therefore$ take $m = 2$

27

Example 9.3 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		

28

Example 9.3
Factorization

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

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

29

Example 9.4
Stock Price Data

- Weekly rates of return for five stocks
  - X_1: J P Morgan
  - X_2: Citibank
  - X_3: Wells Fargo
  - X_4: Royal Dutch Shell
  - X_5: ExxonMobil

30

Example 9.4
Stock Price Data

$$\bar{\mathbf{x}}' = [0.0011 \quad 0.0007 \quad 0.0016 \quad 0.0040 \quad 0.0040]$$

$$\mathbf{R} = \begin{bmatrix} 1 & & & & \\ 0.632 & 1 & & & \\ 0.511 & 0.574 & 1 & & \\ 0.115 & 0.332 & 0.183 & 1 & \\ 0.155 & 0.213 & 0.146 & 0.683 & 1 \end{bmatrix}$$

$$\hat{\lambda}_1 = 2.437, \quad \hat{\mathbf{e}}_1' = [0.469 \quad 0.532 \quad 0.465 \quad 0.387 \quad 0.361]$$

$$\hat{\lambda}_2 = 1.407, \quad \hat{\mathbf{e}}_2' = [-0.368 \quad -0.236 \quad -0.315 \quad 0.585 \quad 0.606]$$

$$\hat{\lambda}_3 = 0.501, \quad \hat{\mathbf{e}}_3' = [-0.604 \quad -0.136 \quad 0.772 \quad 0.093 \quad -0.109]$$

$$\hat{\lambda}_4 = 0.400, \quad \hat{\mathbf{e}}_4' = [0.363 \quad -0.629 \quad 0.289 \quad -0.381 \quad 0.493]$$

$$\hat{\lambda}_5 = 0.255, \quad \hat{\mathbf{e}}_5' = [0.384 \quad -0.496 \quad 0.071 \quad 0.595 \quad -0.498]$$

31

Example 9.4
Principal Component Solution

	One-factor solution		Two-factor solution		
	Estimated factor loadings	Specific variances	Estimated factor loadings		Specific variances
Variable	F_1	$\tilde{\psi}_i = 1 - \tilde{h}_i^2$	F_1	F_2	$\tilde{\psi}_i = 1 - \tilde{h}_i^2$
1. J P Morgan	.732	.46	.732	−.437	.27
2. Citibank	.831	.31	.831	−.280	.23
3. Wells Fargo	.726	.47	.726	−.374	.33
4. Royal Dutch Shell	.605	.63	.605	.694	.15
5. ExxonMobil	.563	.68	.563	.719	.17
Cumulative proportion of total (standardized) sample variance explained	.487		.487	.769	

32

Example 9.4
Residual Matrix for m=2

$$\mathbf{R}-\tilde{\mathbf{L}}\tilde{\mathbf{L}}'-\tilde{\mathbf{\Psi}}$$

$$=\begin{bmatrix} 0 & -0.099 & -0.185 & -0.025 & 0.056 \\ & 0 & -0.134 & 0.014 & -0.054 \\ & & 0 & 0.003 & 0.006 \\ & & & 0 & -0.156 \\ & & & & 0 \end{bmatrix}$$

Maximum Likelihood Method

Assume that $\mathbf{F}_j$ and $\boldsymbol{\varepsilon}_j$ are jointly normal

$\mathbf{X}_j-\boldsymbol{\mu}=\mathbf{L}\mathbf{F}_j+\boldsymbol{\varepsilon}_j$ are then normal, the likelihood is

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

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

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

Uniqueness condition

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

Result 9.1

$\mathbf{X}_1,\mathbf{X}_2,\cdots,\mathbf{X}_n$: random sample from $N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$, $\boldsymbol{\Sigma}=\mathbf{L}\mathbf{L}'+\mathbf{\Psi}$

The maximum likelihood estimators $\hat{\mathbf{L}}$, $\hat{\mathbf{\Psi}}$, and $\hat{\boldsymbol{\mu}}=\bar{\mathbf{x}}$ maximize the likelihood function subject to $\hat{\mathbf{L}}'\hat{\mathbf{\Psi}}^{-1}\hat{\mathbf{L}}$ being diagonal

The maximum likelihood estimates of the communalities are

$$\hat{h}_i^2=\hat{\ell}_{i1}^2+\hat{\ell}_{i2}^2+\cdots+\hat{\ell}_{im}^2$$

so

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

Factorization of R

standardized variable: $\mathbf{Z}=\mathbf{V}^{-1/2}(\mathbf{X}-\boldsymbol{\mu})$

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

$$=\mathbf{L}_\mathbf{z}\mathbf{L}_\mathbf{z}'+\mathbf{\Psi}_\mathbf{z}$$

maximum likelihood estimator

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

$$=\hat{\mathbf{L}}_\mathbf{z}\hat{\mathbf{L}}_\mathbf{z}'+\hat{\mathbf{\Psi}}_\mathbf{z}$$

$$\hat{h}_i^2=\hat{\ell}_{i1}^2+\hat{\ell}_{i2}^2+\cdots+\hat{\ell}_{im}^2$$

$$\begin{pmatrix}\text{Proportion of total (standardized)}\\ \text{sample variance due to the } j\text{th factor}\end{pmatrix}=\frac{\hat{\ell}_{1j}^2+\hat{\ell}_{2j}^2+\cdots+\hat{\ell}_{pj}^2}{p}$$

Example 9.5: Factorization of Stock Price Data

	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	

Example 9.5 ML Residual Matrix

$$\mathbf{R} - \hat{\mathbf{L}}\hat{\mathbf{L}}' - \hat{\mathbf{\Psi}}$$

$$= \begin{bmatrix} 0 & 0.001 & -0.002 & 0.000 & 0.052 \\ & 0 & 0.002 & 0.000 & -0.033 \\ & & 0 & 0.000 & 0.001 \\ & & & 0 & 0.000 \\ & & & & 0 \end{bmatrix}$$

Example 9.6 Olympic Decathlon Data

$$\mathbf{R} = \begin{bmatrix} 1.000 & .6386 & .4752 & .3227 & .5520 & .3262 & .3509 & .4008 & .1821 & -.0352 \\ .6386 & 1.0000 & .4953 & .5668 & .4706 & .3520 & .3998 & .5167 & .3102 & .1012 \\ .4752 & .4953 & 1.0000 & .4357 & .2539 & .2812 & .7926 & .4728 & .4682 & -.0120 \\ .3227 & .5668 & .4357 & 1.0000 & .3449 & .3503 & .3657 & .6040 & .2344 & .2380 \\ .5520 & .4706 & .2539 & .3449 & 1.0000 & .1546 & .2100 & .4213 & .2116 & .4125 \\ .3262 & .3520 & .2812 & .3503 & .1546 & 1.0000 & .2553 & .4163 & .1712 & .0002 \\ .3509 & .3998 & .7926 & .3657 & .2100 & .2553 & 1.0000 & .4036 & .4179 & .0109 \\ .4008 & .5167 & .4728 & .6040 & .4213 & .4163 & .4036 & 1.0000 & .3151 & .2395 \\ .1821 & .3102 & .4682 & .2344 & .2116 & .1712 & .4179 & .3151 & 1.0000 & .0983 \\ -.0352 & .1012 & -.0120 & .2380 & .4125 & .0002 & .0109 & .2395 & .0983 & 1.0000 \end{bmatrix}$$

Example 9.6 Factorization

	Principal component					Maximum likelihood				
	Estimated factor loadings				Specific variances	Estimated factor loadings				Specific variances
Variable	F_1	F_2	F_3	F_4	$\tilde{\psi}_i = 1 - \tilde{h}_i^2$	F_1	F_2	F_3	F_4	$\hat{\psi}_i = 1 - \hat{h}_i^2$
1. 100-m run	.696	.022	−.468	−.416	.12	.993	−.069	−.021	.002	.01
2. Long jump	.793	.075	−.255	−.115	.29	.665	.252	.239	.220	.39
3. Shot put	.771	−.434	.197	−.112	.17	.530	.777	−.141	−.079	.09
4. High jump	.711	.181	.005	.367	.33	.363	.428	.421	.424	.33
5. 400-m run	.605	.549	−.045	−.397	.17	.571	.019	.620	−.305	.20
6. 100 m hurdles	.513	−.083	−.372	.561	.28	.343	.189	.090	.323	.73
7. Discus	.690	−.456	.289	−.078	.23	.402	.718	−.102	−.095	.30
8. Pole vault	.761	.162	.018	.304	.30	.440	.407	.390	.263	.42
9. Javelin	.518	−.252	.519	−.074	.39	.218	.461	.084	−.085	.73
10. 1500-m run	.220	.746	.493	.085	.15	−.016	.091	.609	−.145	.60
Cumulative proportion of total variance explained	.42	.56	.67	.76		.27	.45	.57	.62	

Example 9.6
PC Residual Matrix

$$
\mathbf{R}-\tilde{\mathbf{L}}\tilde{\mathbf{L}}'-\tilde{\boldsymbol{\Psi}}=
\begin{bmatrix}
0 & -.082 & -.006 & -.021 & -.068 & .031 & -.016 & .003 & .039 & .062\\
-.082 & 0 & -.046 & .033 & -.107 & -.078 & -.048 & -.059 & .042 & .006\\
-.006 & -.046 & 0 & .006 & -.010 & -.014 & -.003 & -.013 & -.151 & .055\\
-.021 & .033 & .006 & 0 & -.038 & -.204 & -.015 & -.078 & -.064 & -.086\\
-.068 & -.107 & -.010 & -.038 & 0 & .096 & .025 & -.006 & .030 & -.074\\
.031 & -.078 & -.014 & -.204 & .096 & 0 & .015 & -.124 & .119 & .085\\
-.016 & -.048 & -.003 & -.015 & .025 & .015 & 0 & -.029 & -.210 & .064\\
.003 & -.059 & -.013 & -.078 & -.006 & -.124 & -.029 & 0 & -.026 & -.084\\
.039 & .042 & -.151 & -.064 & .030 & .119 & -.210 & -.026 & 0 & -.078\\
.062 & .006 & .055 & -.086 & -.074 & .085 & .064 & -.084 & -.078 & 0
\end{bmatrix}
$$

41

Example 9.6
ML Residual Matrix

$$
\mathbf{R}-\hat{\mathbf{L}}\hat{\mathbf{L}}'-\hat{\boldsymbol{\Psi}}=
\begin{bmatrix}
0 & .000 & .000 & .-.000 & -.000 & .000 & -.000 & .000 & -.001 & 000\\
.000 & 0 & -.002 & .023 & .005 & .017 & -.003 & -.030 & .047 & -.024\\
.000 & -.002 & 0 & .004 & -.000 & -.009 & .000 & -.001 & -.001 & .000\\
-.000 & .023 & .004 & 0 & -.002 & -.030 & -.004 & -.006 & -.042 & .010\\
-.000 & .005 & -.001 & -.002 & 0 & -.002 & .001 & .001 & .000 & -.001\\
.000 & -.017 & -.009 & -.030 & -.002 & 0 & .022 & .069 & .029 & -.019\\
-.000 & -.003 & .000 & -.004 & .001 & .022 & 0 & -.000 & -.000 & .000\\
.000 & -.030 & -.001 & -.006 & .001 & .069 & -.000 & 0 & .021 & .011\\
-.001 & .047 & -.001 & -.042 & .001 & .029 & -.000 & .021 & 0 & -.003\\
.000 & -.024 & .000 & .010 & -.001 & -.019 & .000 & .011 & -.003 & 0
\end{bmatrix}
$$

42

A Large Sample Test for Number of Common Factors

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

$H_1:\ \boldsymbol{\Sigma}$ any other positive definite matrix

Maximum likelihood function for general case:

proportional to $|\mathbf{S}_n|^{-n/2}e^{-np/2},\ \mathbf{S}_n=(n-1)\mathbf{S}/n$

Maximum likelihood function under $H_0:\hat{\boldsymbol{\Sigma}}=\hat{\mathbf{L}}\hat{\mathbf{L}}'+\hat{\boldsymbol{\Psi}}$

proportional to $|\hat{\mathbf{L}}\hat{\mathbf{L}}'+\hat{\boldsymbol{\Psi}}|^{-n/2}\exp\left(-\frac{n}{2}\operatorname{tr}\left[\left(\hat{\mathbf{L}}\hat{\mathbf{L}}'+\hat{\boldsymbol{\Psi}}\right)^{-1}\mathbf{S}_n\right]\right)$

likelihood ratio statistic for testimg H_0

$$-2\ln\Lambda=-2\ln\left[\frac{\text{maximized likelihood under } H_0}{\text{maximized likelihood}}\right]$$

$$=-2\ln\left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\mathbf{S}_n|}\right)^{-n/2}+n\left[\operatorname{tr}\left(\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S}_n\right)-p\right]$$

43

A Large Sample Test for Number of Common Factors

$$\operatorname{tr}\left(\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S}_n\right)=0\Rightarrow -2\ln\Lambda=n\ln\left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\mathbf{S}_n|}\right)$$

Barlett's chi-square approximation

reject H_0 at the α level of significance if

$$(n-1-(2p+4m+5)/6)\ln\frac{|\hat{\mathbf{L}}\hat{\mathbf{L}}'+\hat{\boldsymbol{\Psi}}|}{|\mathbf{S}_n|}>\chi^2_{[(p-m)^2-p-m]/2}(\alpha)$$

provided that n and $n-p$ are large and

$$m<\frac{1}{2}(2p+1-\sqrt{8p+1})$$

44

Example 9.7
Stock Price Model Testing

$$\frac{|\hat{\mathbf{\Sigma}}|}{|\mathbf{S}_n|} = \frac{|\hat{\mathbf{V}}^{-1/2}|\,|\hat{\mathbf{L}}\hat{\mathbf{L}}'+\hat{\mathbf{\Psi}}|\,|\hat{\mathbf{V}}^{-1/2}|}{|\hat{\mathbf{V}}^{-1/2}|\,|\mathbf{S}_n|\,|\hat{\mathbf{V}}^{-1/2}|} = \frac{|\hat{\mathbf{V}}^{-1/2}\hat{\mathbf{L}}\hat{\mathbf{L}}'\hat{\mathbf{V}}^{-1/2}+\hat{\mathbf{V}}^{-1/2}\hat{\mathbf{\Psi}}\hat{\mathbf{V}}^{-1/2}|}{|\hat{\mathbf{V}}^{-1/2}\mathbf{S}_n\hat{\mathbf{V}}^{-1/2}|}$$

$$= \frac{|\hat{\mathbf{L}}_z\hat{\mathbf{L}}_z' + \hat{\mathbf{\Psi}}_z|}{|\mathbf{R}|} = \frac{0.17898}{0.17519} = 1.0216$$

$$[n-1-(2p+4m+5)/6]\ln\frac{|\hat{\mathbf{L}}\hat{\mathbf{L}}'+\hat{\mathbf{\Psi}}|}{|\mathbf{S}_n|}$$

$$= [103-1-(10+8+5)/6]\ln(1.0216) = 2.10$$

$$\chi^2_{[(p-m)^2-p-m]/2}(0.05) = \chi^2_1(0.05) = 3.84$$

fail to reject H_0 (two-factor model)

45

Outline

- Introduction
- The Orthogonal Factor Model
- Methods of Estimation
- <u>Factor Rotation</u>
- Factor Scores
- Perspectives and Strategy for Factor Analysis

Questions

- What is the factor rotation?
- Why is the factor rotation important for interpretation of the factors?
- Will the communality change under the factor rotation?
- What is the varimax criterion?

Example 9.8
Examination Scores

$\mathbf{R} =$

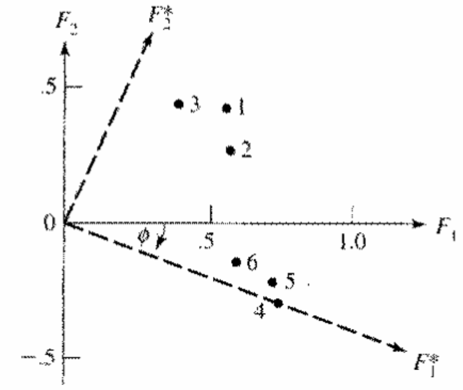
	Gaelic	English	History	Arithmetic	Algebra	Geometry
	1.0	.439	.410	.288	.329	.248
		1.0	.351	.354	.320	.329
			1.0	.164	.190	.181
				1.0	.595	.470
					1.0	.464
						1.0

48

Example 9.8
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 9.8
Factor Rotation



Example 9.8
Rotated Factor Loading

Variable	Estimated rotated factor loadings F_1^*	F_2^*	Communalities $\hat{h}_i^{*2} = \hat{h}_i^2$
1. Gaelic	.369	.594	.490
2. English	.433	.467	.406
3. History	.211	.558	.356
4. Arithmetic	.789	.001	.623
5. Algebra	.752	.054	.568
6. Geometry	.604	.083	.372

Varimax Criterion

$\hat{\mathbf{L}}^* = \hat{\mathbf{L}}\mathbf{T}, \quad \mathbf{TT'} = \mathbf{I}, \quad \tilde{\ell}_{ij}^* = \hat{\ell}_{ij}^* / \hat{h}_i$

select **T** that makes

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

as large as possible

Interpretation :

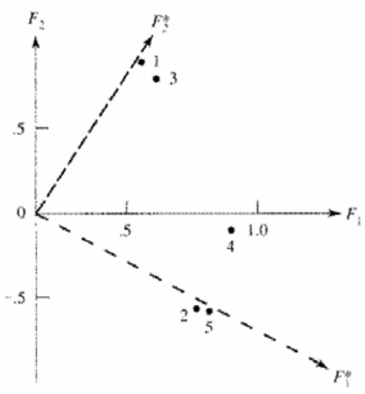
$$V \propto \sum_{j=1}^{m} \left(\begin{array}{l} \text{variance of squares of (scaled)} \\ \text{loadings for } j\text{th factor} \end{array} \right)$$

i.e., "spreading out" the squares of the loadings on each factor as much as possible

Example 9.9: Consumer-Preference Factor Analysis

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

Example 9.9 Factor Rotation



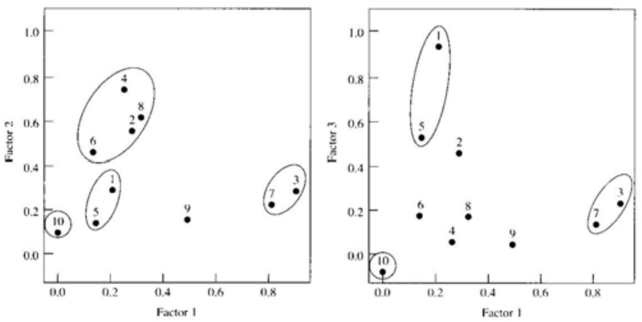
Example 9.10 Stock Price Factor Analysis

Variable	Maximum likelihood estimates of factor loadings F_1	F_2	Rotated estimated factor loadings F_1^*	F_2^*	Specific variances $\hat{\psi}_i^2 = 1 - \hat{h}_i^2$
J P Morgan	.115	.755	.763	.024	.42
Citibank	.322	.788	.821	.227	.27
Wells Fargo	.182	.652	.669	.104	.54
Royal Dutch Shell	1.000	−.000	.118	.993	.00
ExxonMobil	.683	.032	.113	.675	.53
Cumulative proportion of total sample variance explained	.323	.647	.346	.647	

Example 9.11 Olympic Decathlon Factor Analysis

	Principal component					Maximum likelihood				
	Estimated rotated factor loadings, $\tilde{\ell}_{ij}^*$				Specific variances	Estimated rotated factor loadings, $\hat{\ell}_{ij}^*$				Specific variances
Variable	F_1^*	F_2^*	F_3^*	F_4^*	$\tilde{\psi}_i = 1 - \tilde{h}_i^2$	F_1^*	F_2^*	F_3^*	F_4^*	$\hat{\psi}_i = 1 - \hat{h}_i^2$
100-m run	.182	.885	.205	−.139	.12	.204	.296	.928	−.005	.01
Long jump	.291	.664	.429	.055	.29	.280	.554	.451	.155	.39
Shot put	.819	.302	.252	−.097	.17	.883	.278	.228	−.045	.09
High jump	.267	.221	.683	.293	.33	.254	.739	.057	.242	.33
400-m run	.086	.747	.068	.507	.17	.142	.151	.519	.700	.20
110-m hurdles	.048	.108	.826	−.161	.28	.136	.465	.173	−.033	.73
Discus	.832	.185	.204	−.076	.23	.793	.220	.133	−.009	.30
Pole vault	.324	.278	.656	.293	.30	.314	.613	.169	.279	.42
Javelin	.754	.024	.054	.188	.39	.477	.160	.041	.139	.73
1500-m run	−.002	.019	.075	.921	.15	.001	.110	−.070	.619	.60
Cumulative proportion of total sample variance explained	.22	.43	.62	.76		.20	.37	.51	.62	

Example 9.11
Rotated ML Loadings



57

Outline

- Introduction
- The Orthogonal Factor Model
- Methods of Estimation
- Factor Rotation
- Factor Scores
- Perspectives and Strategy for Factor Analysis

Questions

- What is the factor scores?
- How to solve the factor scores by the weighted least squares method?
- How to solve the factor scores by the principal component analysis?
- How to solve the factor scores by the regression model?

Factor Scores

$\hat{\mathbf{f}}_j$ = estimate of the values $\mathbf{f}_j$ attained by $\mathbf{F}_j$ (jth case)

used for diagnostic purposes and inputs to next analysis

Essences of our estimations

1. Treat $\hat{\ell}_{ij}$ and $\hat{\psi}_i$ as the true values
2. Involve linear transformations of the original data

60

Weighted Least Squares Method

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

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

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

use maximul likelihood estimates

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

for correlation matrix

$\hat{\mathbf{f}}_j=(\hat{\mathbf{L}}_{\mathbf{z}}'\hat{\boldsymbol{\Psi}}_{\mathbf{z}}^{-1}\hat{\mathbf{L}}_{\mathbf{z}})^{-1}\hat{\mathbf{L}}_{\mathbf{z}}'\hat{\boldsymbol{\Psi}}_{\mathbf{z}}^{-1}\mathbf{z}_j=\hat{\boldsymbol{\Delta}}_{\mathbf{z}}^{-1}\hat{\mathbf{L}}_{\mathbf{z}}'\hat{\boldsymbol{\Psi}}_{\mathbf{z}}^{-1}\mathbf{z}_j$

$\mathbf{z}_j=\mathbf{D}^{-1/2}(\mathbf{x}_j-\overline{\mathbf{x}}),\quad \hat{\boldsymbol{\rho}}=\hat{\mathbf{L}}_{\mathbf{z}}\hat{\mathbf{L}}_{\mathbf{z}}'+\hat{\boldsymbol{\Psi}}_{\mathbf{z}}$

61

Factor Scores of Principal Component Method

Use unweighted least square method (ψ_i are nearly equal)

$\hat{\mathbf{f}}_j=\left(\tilde{\mathbf{L}}'\tilde{\mathbf{L}}\right)^{-1}\tilde{\mathbf{L}}'\left(\mathbf{x}_j-\overline{\mathbf{x}}\right)\text{or }\hat{\mathbf{f}}_j=\left(\tilde{\mathbf{L}}_{\mathbf{z}}'\tilde{\mathbf{L}}_{\mathbf{z}}\right)^{-1}\tilde{\mathbf{L}}_{\mathbf{z}}'\mathbf{z}_j$

$\tilde{\mathbf{L}}=\left[\sqrt{\hat{\lambda}_1}\hat{\mathbf{e}}_1\quad\sqrt{\hat{\lambda}_2}\hat{\mathbf{e}}_2\quad\cdots\quad\sqrt{\hat{\lambda}_m}\hat{\mathbf{e}}_m\right]$

$$\hat{\mathbf{f}}_j=\begin{bmatrix}\frac{1}{\sqrt{\hat{\lambda}_1}}\hat{\mathbf{e}}_1'\left(\mathbf{x}_j-\overline{\mathbf{x}}\right)\\ \frac{1}{\sqrt{\hat{\lambda}_2}}\hat{\mathbf{e}}_2'\left(\mathbf{x}_j-\overline{\mathbf{x}}\right)\\ \vdots\\ \frac{1}{\sqrt{\hat{\lambda}_m}}\hat{\mathbf{e}}_m'\left(\mathbf{x}_j-\overline{\mathbf{x}}\right)\end{bmatrix},\quad \frac{1}{n}\sum_{j=1}^{n}\hat{\mathbf{f}}_j=0,\quad \frac{1}{n-1}\sum_{j=1}^{n}\hat{\mathbf{f}}_j\hat{\mathbf{f}}_j'=\mathbf{I}$$

62

Orthogonal Factor Model

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

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

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

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

63

Regression Model

$\mathbf{X}-\boldsymbol{\mu}=\mathbf{LF}+\boldsymbol{\varepsilon}: N_p(\mathbf{0},\mathbf{LL}'+\boldsymbol{\Psi})$

$\mathbf{X}-\boldsymbol{\mu}$ and $\mathbf{F}$ joint normal: $N_{p+m}(\mathbf{0},\boldsymbol{\Sigma}^*)$

$$\boldsymbol{\Sigma}^*=\begin{bmatrix}\boldsymbol{\Sigma}=\mathbf{LL}'+\boldsymbol{\Psi} & \mathbf{L}\\ \mathbf{L}' & \mathbf{I}\end{bmatrix}$$

$E(\mathbf{F}\,|\,\mathbf{x})=\mathbf{L}'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})=\mathbf{L}'(\mathbf{LL}'+\boldsymbol{\Psi})^{-1}(\mathbf{x}-\boldsymbol{\mu})$

$\text{Cov}(\mathbf{F}\,|\,\mathbf{x})=\mathbf{I}-\mathbf{L}'\boldsymbol{\Sigma}^{-1}\mathbf{L}=\mathbf{I}-\mathbf{L}'(\mathbf{LL}'+\boldsymbol{\Psi})^{-1}\mathbf{L}$

$\hat{\mathbf{f}}_j=\hat{\mathbf{L}}'(\hat{\mathbf{L}}\hat{\mathbf{L}}'+\hat{\boldsymbol{\Psi}})^{-1}(\mathbf{x}_j-\overline{\mathbf{x}}),\quad \hat{\mathbf{L}}'(\hat{\mathbf{L}}\hat{\mathbf{L}}'+\hat{\boldsymbol{\Psi}})^{-1}=(\mathbf{I}+\hat{\mathbf{L}}'\hat{\boldsymbol{\Psi}}^{-1}\hat{\mathbf{L}})^{-1}\hat{\mathbf{L}}'\hat{\boldsymbol{\Psi}}^{-1}$

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

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

64

Factor Scores by Regression

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

or

$$\hat{\mathbf{f}}_j = \hat{\mathbf{L}}_{\mathbf{z}}'\mathbf{R}^{-1}\mathbf{z}_j$$

$$\mathbf{z}_j = \mathbf{D}^{-1/2}(\mathbf{x}_j - \overline{\mathbf{x}}), \quad \hat{\boldsymbol{\rho}} = \hat{\mathbf{L}}_{\mathbf{z}}\hat{\mathbf{L}}_{\mathbf{z}}' + \hat{\boldsymbol{\Psi}}_{\mathbf{z}}$$

65

Example 9.12 Stock Price Data

maximum likelihood solution from $\mathbf{R}$

$$\hat{\mathbf{L}}_{\mathbf{z}}^* = \begin{bmatrix} 0.763 & 0.024 \\ 0.821 & 0.227 \\ 0.669 & 0.104 \\ 0.118 & 0.993 \\ 0.113 & 0.675 \end{bmatrix}, \hat{\boldsymbol{\Psi}}_{\mathbf{z}} = \begin{bmatrix} 0.42 & 0 & 0 & 0 & 0 \\ 0 & 0.27 & 0 & 0 & 0 \\ 0 & 0 & 0.54 & 0 & 0 \\ 0 & 0 & 0 & 0.00 & 0 \\ 0 & 0 & 0 & 0 & 0.53 \end{bmatrix}$$

$$\mathbf{z}' = \begin{bmatrix} 0.50 & -1.40 & -0.20 & -0.70 & 1.40 \end{bmatrix}$$

weighted least squares

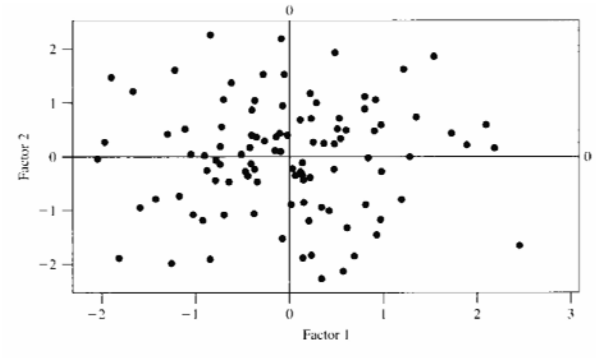
$$\hat{\mathbf{f}} = \left(\hat{\mathbf{L}}_{\mathbf{z}}^{*\prime}\hat{\boldsymbol{\Psi}}_{\mathbf{z}}^{-1}\hat{\mathbf{L}}_{\mathbf{z}}^*\right)^{-1}\hat{\mathbf{L}}_{\mathbf{z}}^{*\prime}\hat{\boldsymbol{\Psi}}_{\mathbf{z}}^{-1}\mathbf{z} = \begin{bmatrix} -0.61 & -0.61 \end{bmatrix}'$$

regression

$$\hat{\mathbf{f}} = \hat{\mathbf{L}}_{\mathbf{z}}^{*\prime}\mathbf{R}^{-1}\mathbf{z} = \begin{bmatrix} -0.5 & -0.64 \end{bmatrix}'$$

66

Example 9.12 Factor Scores by Regression



67

Example 9.13: Simple Summary Scores for Stock Price Data

principal component factor loading

$$\tilde{\mathbf{L}} = \begin{bmatrix} 0.732 & -0.437 \\ 0.831 & -0.280 \\ 0.726 & -0.374 \\ 0.605 & 0.694 \\ 0.563 & 0.719 \end{bmatrix}, \quad \tilde{\mathbf{L}}^* = \tilde{\mathbf{L}}\mathbf{T} = \begin{bmatrix} 0.852 & 0.030 \\ 0.851 & 0.214 \\ 0.813 & 0.079 \\ 0.133 & 0.911 \\ 0.084 & 0.909 \end{bmatrix}$$

summary scores

$$\hat{f}_1 = x_1 + x_2 + x_3 + x_4 + x_5, \ \hat{f}_2 = x_4 + x_5 - x_1 \quad (\tilde{\mathbf{L}})$$

$$\hat{f}_1 = x_1 + x_2 + x_3, \ \hat{f}_2 = x_4 + x_5 \quad (\tilde{\mathbf{L}}^*)$$

68

Outline

- Introduction
- The Orthogonal Factor Model
- Methods of Estimation
- Factor Rotation
- Factor Scores
- Perspectives and Strategy for Factor Analysis

Questions

- Give a strategy for factor analysis?
- What is the “Wow” criterion?

A Strategy for Factor Analysis

- 1. Perform a principal component factor analysis
  - Look for suspicious observations by plotting the factor scores
  - Try a varimax rotation
- 2. Perform a maximum likelihood factor analysis, including a varimax rotation

71

A Strategy for Factor Analysis

- 3. Compare the solutions obtained from the two factor analyses
  - Do the loadings group in the same manner?
  - Plot factor scores obtained for PC against scores from ML analysis
- 4. Repeat the first 3 steps for other numbers of common factors
- 5. For large data sets, split them in half and perform factor analysis on each part. Compare the two results with each other and with that from the complete data set

72

Example 9.14 Chicken-Bone Data

$n = 276$ measurements on bone dimensions

Head : X_1 = skull length, X_2 = skull breadth

Leg : X_3 = femur length, X_4 = tibia length

Wing : X_5 = humerus length, X_6 = ulna length

$$\mathbf{R} = \begin{bmatrix} 1 & 0.505 & 0.569 & 0.602 & 0.621 & 0.603 \\ & 1 & 0.422 & 0.467 & 0.482 & 0.450 \\ & & 1 & 0.926 & 0.877 & 0.878 \\ & & & 1 & 0.874 & 0.894 \\ & & & & 1 & 0.937 \\ & & & & & 1 \end{bmatrix}$$

73

Example 9.14:Principal Component Factor Analysis Results

Variable	Estimated factor loadings F_1	F_2	F_3	Rotated estimated loadings F_1^*	F_2^*	F_3^*	$\tilde{\psi}_i$
1. Skull length	.741	.350	.573	.355	.244	.902	.00
2. Skull breadth	.604	.720	−.340	.235	.949	.211	.00
3. Femur length	.929	−.233	−.075	.921	.164	.218	.08
4. Tibia length	.943	−.175	−.067	.904	.212	.252	.08
5. Humerus length	.948	−.143	−.045	.888	.228	.283	.08
6. Ulna length	.945	−.189	−.047	.908	.192	.264	.07
Cumulative proportion of total (standardized) sample variance explained	.743	.873	.950	.576	.763	.950	

74

Example 9.14: Maximum Likelihood Factor Analysis Results

Variable	Estimated factor loadings F_1	F_2	F_3	Rotated estimated loadings F_1^*	F_2^*	F_3^*	$\hat{\psi}_i$
1. Skull length	.602	.214	.286	.467	.506	.128	.51
2. Skull breadth	.467	.177	.652	.211	.792	.050	.33
3. Femur length	.926	.145	−.057	.890	.289	.084	.12
4. Tibia length	1.000	.000	−.000	.936	.345	−.073	.00
5. Humerus length	.874	.463	−.012	.831	.362	.396	.02
6. Ulna length	.894	.336	−.039	.857	.325	.272	.09
Cumulative proportion of total (standardized) sample variance explained	.667	.738	.823	.559	.779	.823	

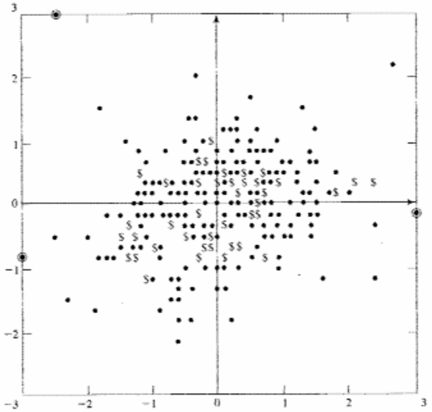
75

Example 9.14 Residual Matrix for ML Estimates

$$\mathbf{R} - \hat{\mathbf{L}}_z \hat{\mathbf{L}}_z' - \hat{\boldsymbol{\Psi}}_z = \begin{bmatrix} 0.000 & & & & & \\ -0.000 & 0.000 & & & & \\ -0.003 & 0.001 & 0.000 & & & \\ 0.000 & 0.000 & 0.000 & 0.000 & & \\ -0.001 & 0.000 & 0.000 & 0.000 & 0.000 & \\ 0.004 & -0.001 & -0.001 & 0.000 & -0.000 & 0.000 \end{bmatrix}$$

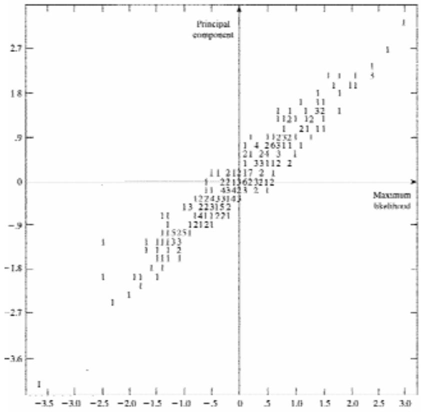
76

Example 9.14
Factor Scores for Factors 1 & 2



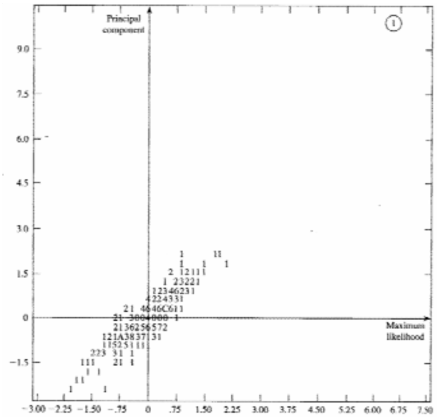
77

Example 9.14
Pairs of Factor Scores: Factor 1



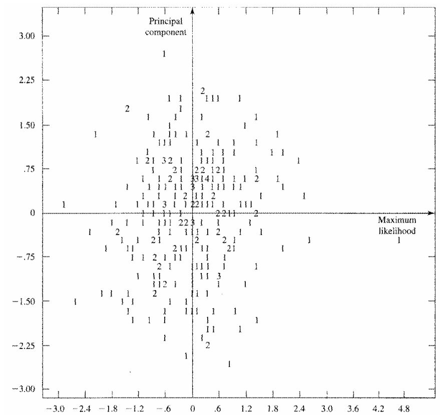
78

Example 9.14
Pairs of Factor Scores: Factor 2



79

Example 9.14
Pairs of Factor Scores: Factor 3



80

Example 9.14 Divided Data Set

$$n_1 = 137, \quad \mathbf{R}_1 = \begin{bmatrix} 1 & & & & & \\ 0.696 & 1 & & & & \\ 0.588 & 0.540 & 1 & & & \\ 0.639 & 0.575 & 0.901 & 1 & & \\ 0.694 & 0.606 & 0.844 & 0.835 & 1 & \\ 0.660 & 0.584 & 0.866 & 0.863 & 0.931 & 1 \end{bmatrix}$$

$$n_2 = 139, \quad \mathbf{R}_2 = \begin{bmatrix} 1 & & & & & \\ 0.366 & 1 & & & & \\ 0.572 & 0.352 & 1 & & & \\ 0.587 & 0.406 & 0.950 & 1 & & \\ 0.587 & 0.420 & 0.909 & 0.911 & 1 & \\ 0.598 & 0.386 & 0.894 & 0.927 & 0.940 & 1 \end{bmatrix}$$

Example 9.14: PC Factor Analysis for Divided Data Set

	First set (n_1 = 137 observations) Rotated estimated factor loadings				Second set (n_2 = 139 observations) Rotated estimated factor loadings			
Variable	F_1^*	F_2^*	F_3^*	$\tilde{\psi}_i$	F_1^*	F_2^*	F_3^*	$\tilde{\psi}_i$
1. Skull length	.360	.361	.853	.01	.352	.921	.167	.00
2. Skull breadth	.303	.899	.312	.00	.203	.145	.968	.00
3. Femur length	.914	.238	.175	.08	.930	.239	.130	.06
4. Tibia length	.877	.270	.242	.10	.925	.248	.187	.05
5. Humerus length	.830	.247	.395	.11	.912	.252	.208	.06
6. Ulna length	.871	.231	.332	.08	.914	.272	.168	.06
Cumulative proportion of total (standardized) sample variance explained	.546	.743	.940		.593	.780	.962	

WOW Criterion

- In practice the vast majority of attempted factor analyses do not yield clear-cut results
- If, while scrutinizing the factor analysis, the investigator can shout "Wow, I understand these factors," the application is deemed successful