多元统计分析

第9讲 因子分析(II)

Johnson & Wichern Ch9.4-9.6

统计学研究中心 邓婉璐
wanludeng@tsinghua.edu.cn

2018-2019春季学期

Outline

- Introduction and Model
- Methods for Estimation
  - PC method
  - MLE method
- Explanation – Rotation
- Factor Scores
  - Weighted LSE Method
  - Regression Method

Rotation for Interpretation

Factor Rotation

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

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

- Question
  - Why is it not unique in FA, but unique in PCA?
  - Is the covariance/correlation matrix changed after rotation? What about residual matrix, estimated specific variances, communalities?
  - Why rotation? (Interpretation)
  - Criteria? What is a desirable result?

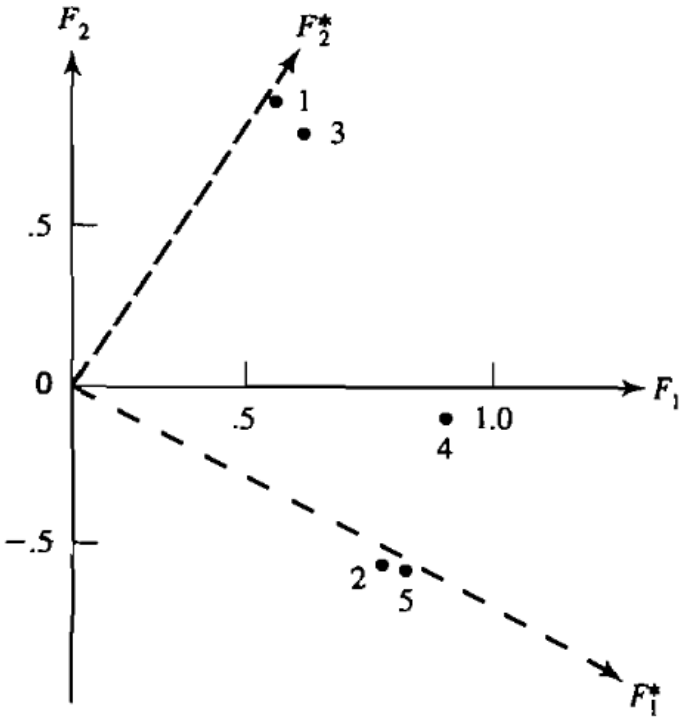
PCA的几何投影：PCA是找全部主成分，使得投影到任意维数，主成分确定的平面都是最优的。而FA只找一个平面，所以当固定维数的时候，可以在该平面内旋转，不影响投影结果，但却有很好的解释度

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.

Variable	Estimated factor loadings F_1	F_2
Taste	1	0
Good buy for money	0	1
Flavor	1	0
Suitable for snack	0	1
Provides lots of energy	0	1

爱好因子（附加） 食物因子（实用）



Varimax Criterion

- Importance depends on magnitude not sign: $(\hat{l}_{ij}^{*})^2$
- Scaling: $(\tilde{l}_{ij}^{*})^2 = (\hat{l}_{ij}^{*})^2 / \hat{h}_i^2$ 物以稀为贵
- Varimax procedure selects the orthogonal transformation **T** that maximizes

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

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

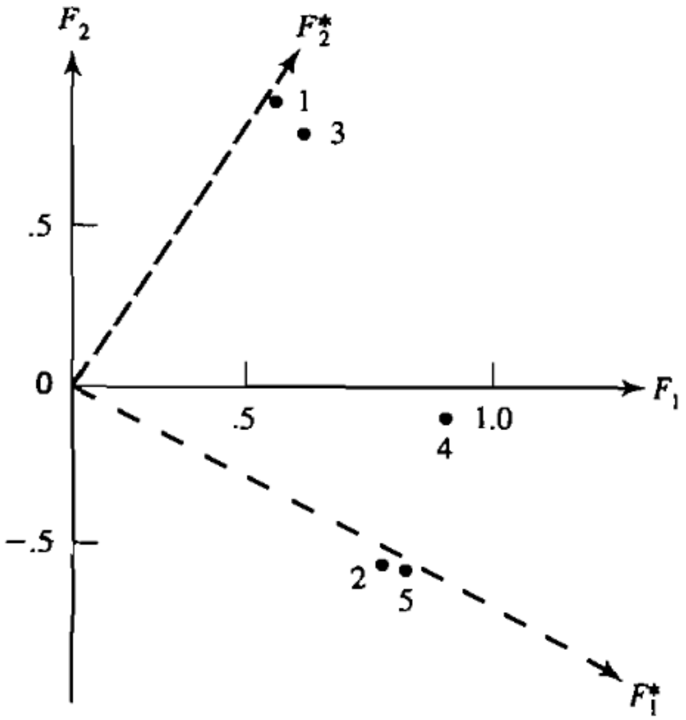
Maximizing V corresponds to ‘spreading out’ the squares of the loadings on each factor as much as possible.

Example: Consumer Preference

- Is the explained variance changed after rotation?

Table 9.7

Variable	Estimated factor loadings F_1	F_2	Rotated estimated factor loadings F_1^*	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	



Estimating Factor Scores

Why Do We Want Factor Score?

- Why we want factor score?
  - used for diagnostic purposes
  - or inputs to a subsequent analysis

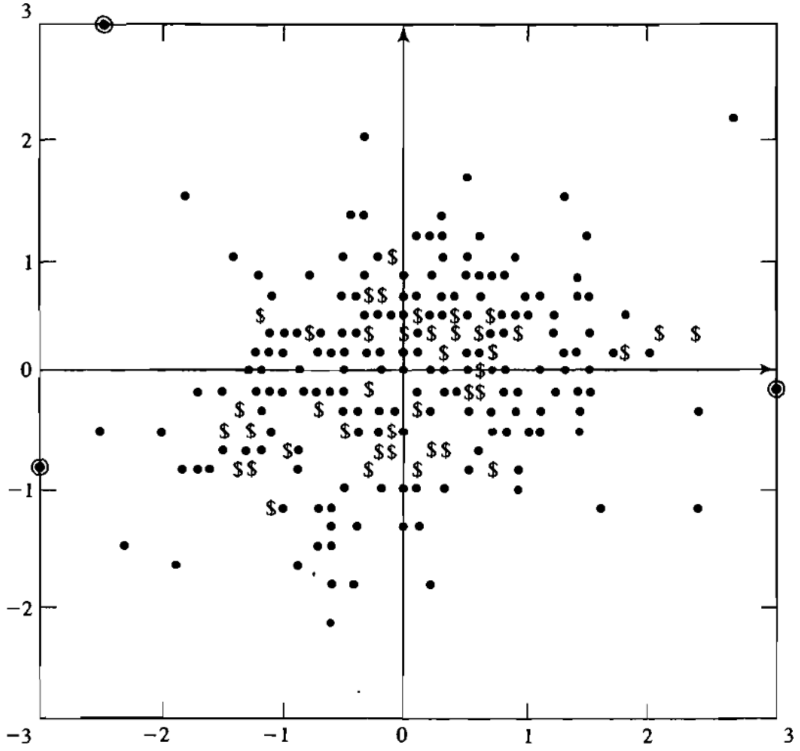
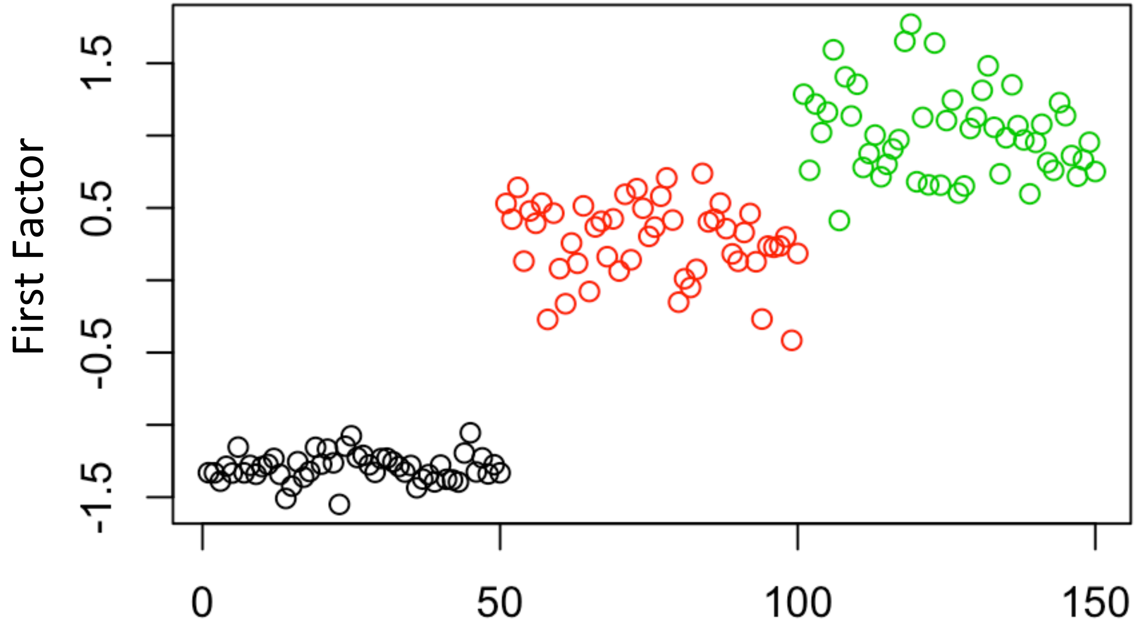


Figure 9.5 Factor scores for the first two factors of chicken-bone data.



Example: Iris data

How to Find Factor Score?

- Mission impossible: too many latent variables

- Solution: Take L, Ψ as known. Get factor scores heuristically.

- Methods for getting factor scores
  - Weighted least squares method
  - Ordinary least squares method (with L, Ψ obtained from PC approach)
  - Regression method

- Comparison
  - WLS *vs* Regression
  - Factor scores *vs* Principle scores

Weighted Least Squares Method

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

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

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

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

Bartlett Score (1954)

Question:

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

Ordinary Least Squares Method

- If we use PC approach to estimate L, Ψ , then we usually choose OLS to obtain the factor score.

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

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

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

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

- Connection to PC scores (y_j): the j^{th} factor score $\underset{(1\times 1)}{\hat{f}_j} = \frac{1}{\sqrt{\lambda_j}} e_j'(x-\mu) = \frac{1}{\sqrt{\lambda_j}} y_j$

Regression Method

- Recall: multivariate normal distribution, conditional distribution

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

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

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

Regression Method

$$\begin{pmatrix} \mathbf{X}-\mu \\ \mathbf{F} \end{pmatrix} = \begin{pmatrix} \mathbf{L} \\ \mathbf{I}_m \end{pmatrix}\mathbf{F} + \begin{pmatrix} \varepsilon \\ \mathbf{0} \end{pmatrix}, \mathbf{F} \sim N_m(\mathbf{0},\mathbf{I})$$

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

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

Relationship to regression analysis

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

The factor score vector for the j^{th} unit is given by

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

Question: What if the correlation matrix is factored? $\hat{\mathbf{f}}_j = \hat{\mathbf{L}}_z'\hat{\mathbf{R}}^{-1}\mathbf{z}_{\mathbf{j}}$

Regression Method – Another Perspective

- Model: $$\underset{(m\times 1)}{F} = \underset{(m\times p)}{B}\underset{(p\times 1)}{(X-\mu)} + \underset{(m\times 1)}{\eta}, E\eta = 0, \operatorname{cov}(X-\mu,\eta) = 0.$$

- Difficulty: F is not observed

- Solution: recall that we take *L* as known, and

$$l_{ij} = \operatorname{cov}(X_i-\mu_i, F_j) = b_{j1}\sigma_{i1} + \cdots + b_{jp}\sigma_{ip} = [\sigma_{i1},\cdots,\sigma_{ip}]\begin{bmatrix} b_{j1} \\ \vdots \\ b_{jp} \end{bmatrix}$$

$$\underset{(p\times m)}{L} = (l_{ij}) = \Sigma B'$$

$$\hat{B} = \hat{\Sigma}^{-1}\hat{L},\ \underset{(m\times 1)}{\hat{F}} = \underset{(m\times p)}{\hat{L}'}\underset{(p\times p)}{\hat{\Sigma}^{-1}}\underset{(p\times 1)}{(X-\mu)}$$

Thompson Score (1939)

Regression Method – Further Notes

- Approximation:

  $\mathbf{S}$ is often used for $\hat{\Sigma}$ rather than $\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\Psi}$.

  Then the factor score vector for the j^{th} unit is $\hat{\mathbf{f}}_j = \hat{\mathbf{L}}'\mathbf{S}^{-1}(\mathbf{x}_\mathbf{j} - \bar{\mathbf{x}})$

- Connection to PC scores:

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

$$\mathbf{S} = \mathbf{P}\Lambda\mathbf{P}'$$

$$\underset{(p\times m)}{\mathbf{L}} = \underset{(p\times m)}{\mathbf{P}_m}\,\underset{(m\times m)}{\Lambda_m^{1/2}} = [e_1, \cdots, e_m]\begin{bmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_m} \end{bmatrix}$$

$$\begin{aligned} \hat{\mathbf{f}}_j &= \hat{\mathbf{L}}'\mathbf{S}^{-1}(\mathbf{x}_\mathbf{j} - \bar{\mathbf{x}}) \\ &= (\mathbf{P}_m\Lambda_m^{1/2})'(\mathbf{P}\Lambda^{-1}\mathbf{P}')(\mathbf{x}_\mathbf{j} - \bar{\mathbf{x}}) \\ &= \Lambda_m^{-1/2}\boxed{\mathbf{P}_m'(\mathbf{x}_\mathbf{j} - \bar{\mathbf{x}})} \end{aligned}$$

PC score

Comparison between two methods

- Estimation

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

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

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

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

Comparison between two methods

- Unbiasedness
  - Estimation under weighted least squares method is unbiased.

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

$$E(\hat{\mathbf{f}}^{WLS} \mid F) = (\mathbf{L}'\Psi^{-1}\mathbf{L})^{-1}\mathbf{L}'\Psi^{-1}\mathbf{L}F = F$$

  - Estimation under regression method is biased.

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

$$E(\hat{\mathrm{f}}^{\mathrm{R}} \mid F) = E(\mathbf{L}'(\mathbf{L}\mathbf{L}'+\Psi)^{-1}(\mathbf{x}-\mu) \mid F) = \mathbf{L}'(\mathbf{L}\mathbf{L}'+\Psi)^{-1}\mathbf{L}F$$

$$= (I + (L'\Psi^{-1}L)^{-1})^{-1} F$$

(by $\hat{\mathbf{L}}'(\hat{\mathbf{L}}\hat{\mathbf{L}}'+\hat{\Psi})^{-1} = (I+\hat{\mathbf{L}}'\hat{\Psi}^{-1}\hat{\mathbf{L}})^{-1}\hat{\mathbf{L}}'\hat{\Psi}^{-1}$)

Can also be derived directly with results on last page.

Comparison between two methods

- Average prediction error
  - For the estimation under weighted least squares method:

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

$$= (\mathbf{L}'\Psi^{-1}\mathbf{L})^{-1}\mathbf{L}'\Psi^{-1}\mathbf{L}F + (\mathbf{L}'\Psi^{-1}\mathbf{L})^{-1}\mathbf{L}'\Psi^{-1}\varepsilon = F + (\mathbf{L}'\Psi^{-1}\mathbf{L})^{-1}\mathbf{L}'\Psi^{-1}\varepsilon$$

$$\boxed{E((\hat{\mathbf{f}}^{WLS}-F)(\hat{\mathbf{f}}^{WLS}-F)')} = (\mathbf{L}'\Psi^{-1}\mathbf{L})^{-1}\mathbf{L}'\Psi^{-1}\operatorname{var}(\varepsilon)\Psi^{-1}\mathbf{L}(\mathbf{L}'\Psi^{-1}\mathbf{L})^{-1}\boxed{= (\mathbf{L}'\Psi^{-1}\mathbf{L})^{-1}}$$

Larger

  - For the estimation under regression method:

$$\hat{\mathrm{f}}^{\mathrm{R}} = \mathbf{L}'(\mathbf{L}\mathbf{L}'+\Psi)^{-1}(\mathbf{x}-\mu) = \mathbf{L}'\Sigma^{-1}(\mathbf{x}-\mu) = \mathbf{L}'\Sigma^{-1}\mathbf{L}F + \mathbf{L}'\Sigma^{-1}\varepsilon$$

$$\boxed{E((\hat{\mathrm{f}}^{\mathrm{R}}-F)(\hat{\mathrm{f}}^{\mathrm{R}}-F)')} = (I+\Delta)^{-1}(I+\Delta)^{-1} + (I+\Delta)^{-1}\Delta(I+\Delta)^{-1}$$

(by $\mathbf{L}'(\mathbf{L}\mathbf{L}'+\Psi)^{-1}\mathbf{L} - I = -(I+\mathbf{L}'\Psi^{-1}\mathbf{L})^{-1}$, and denote $\Delta = \mathbf{L}'\Psi^{-1}\mathbf{L}$)

$$\boxed{= (I + L'\Psi^{-1}L)^{-1}}$$

Smaller

Influence of Factor Rotation on FS

- Theoretically, we expect: $\mathbf{L}^* = \mathbf{L}\mathbf{T}, \mathbf{f}_j^* = \mathbf{T}'\mathbf{f}_j$
- The estimations keep this property: $\hat{\mathbf{L}}^* = \hat{\mathbf{L}}\mathbf{T}, \hat{\mathbf{f}}_j^* = \mathbf{T}'\hat{\mathbf{f}}_j$

$$\hat{\mathrm{f}}_{\mathrm{j}}^{\mathrm{R}} = \hat{\mathbf{L}}'\hat{\Sigma}^{-1}(\mathbf{x}_j - \overline{\mathbf{x}}) = \hat{\mathbf{L}}'(\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\Psi})^{-1}(\mathbf{x}_j - \overline{\mathbf{x}}) \Rightarrow \hat{\mathbf{f}}_j^{\mathrm{R}*} = \mathbf{T}'\hat{\mathbf{f}}_j^{\mathrm{R}}$$

$$\hat{\mathbf{f}}_j^{WLS} = (\hat{\mathbf{L}}'\hat{\Psi}^{-1}\hat{\mathbf{L}})^{-1}\hat{\mathbf{L}}'\hat{\Psi}^{-1}(\mathbf{x}_j - \vec{\mathbf{x}}) \Rightarrow \hat{\mathbf{f}}_j^{WLS*} = \mathbf{T}'\hat{\mathbf{f}}_j^{WLS}$$

R code

MLE method:
with data

```
> res <- factanal(iris[,1:4],1,scores = 'regression',rotate='varimax')
> res

Call:
factanal(x = iris[, 1:4], factors = 1, scores = "regression",     rotat
arimax")

Uniquenesses:
Sepal.Length  Sepal.Width Petal.Length  Petal.Width
       0.240        0.822        0.005        0.069

Loadings:
             Factor1
Sepal.Length  0.872
Sepal.Width  -0.422
Petal.Length  0.998
Petal.Width   0.965

               Factor1
SS loadings      2.864
Proportion Var   0.716

Test of the hypothesis that 1 factor is sufficient.
The chi square statistic is 85.51 on 2 degrees of freedom.
The p-value is 2.7e-19
```

MLE method:
with covariance matrix

```
> R <- var(iris[,1:4])
> res <- factanal(R,1,covmat=R)
```

PC method:
psych::principal

Strategy in Real Application

Strategy for Factor Analysis

- Guide:
  - ✧Interpretation, understanding ('WOW' Rule)
  - ✧Consistency
- Steps:
  1. For large data sets, split them in half and perform FA on each part
  2. For given data and fixed *m*
     (1) Perform a principal component factor analysis
     (2) Perform a maximum likelihood factor analysis
     (3) Compare the solutions obtained from the two factor analyses
  3. Repeat the step 1 for the other number of common factors *m*
  4. Check the consistency and interpretation.

Illustration through Examples

- Data

The full data set consists of $n = 276$ measurements on bone dimensions:

$$\text{Head:} \quad \begin{cases} X_1 = \text{skull length} \\ X_2 = \text{skull breadth} \end{cases}$$

$$\text{Leg:} \quad \begin{cases} X_3 = \text{femur length} \\ X_4 = \text{tibia length} \end{cases}$$

$$\text{Wing:} \quad \begin{cases} X_5 = \text{humerus length} \\ X_6 = \text{ulna length} \end{cases}$$

Illustration through Examples

- Step 2.(1)(2)

Principal Component

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	

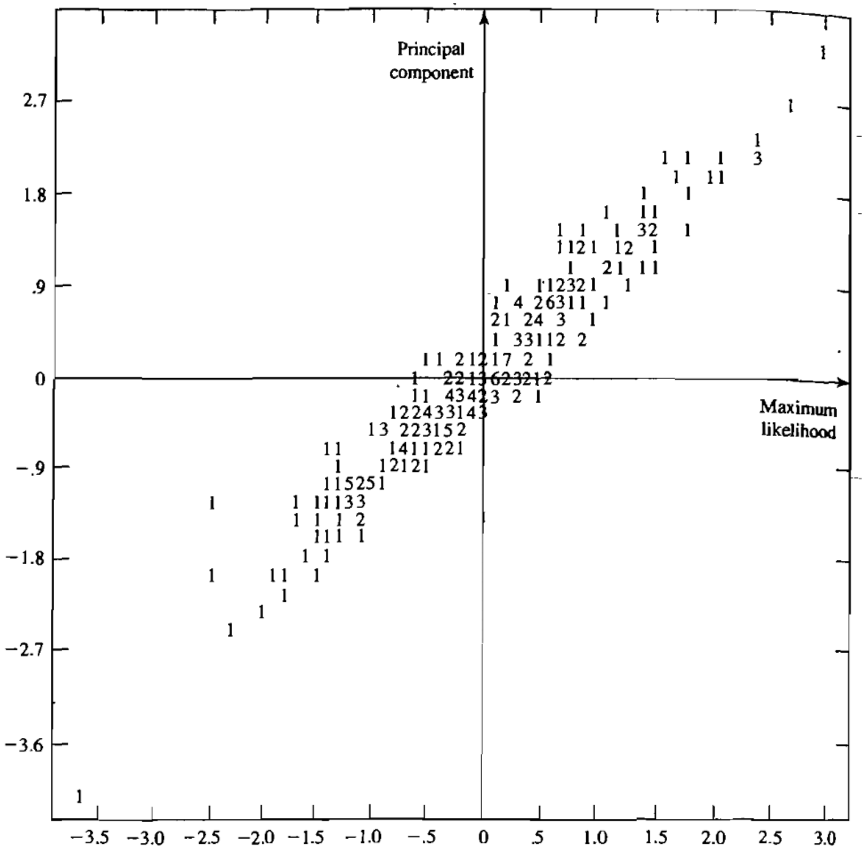
Maximum Likelihood

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	

Illustration through Examples

- Step 2.(3)

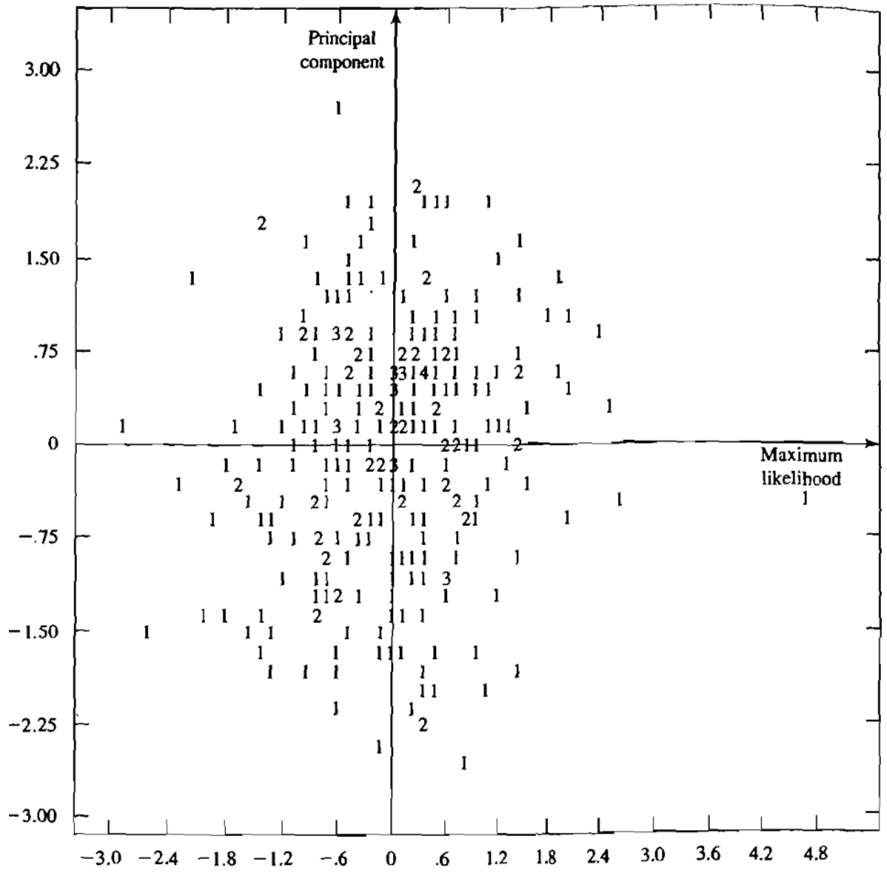
1st Factor Score



(a) First factor

Figure 9.6 Pairs of factor scores for the chicken-bone data. (Loadings are estimated by principal component and maximum likelihood methods.)

3rd Factor Score



(c) Third factor

Figure 9.6 *(continued)*

Summary

Summary

- Factor Rotation
  - Motivation: Interpretation
  - Varimax criteria
- Factor Scores
  - Motivation: diagnostic, further study
  - Weighted least squares method (Bartlett score), OLS (PC estimation for L)
  - Regression method (Thompson score)
  - Comparison for understanding:
    - WLS vs Regression method (trade-off)
    - Covariance matrix vs correlation matrix
    - Factor scores vs Principle scores
- Guide for application