

# Multivariate Statistics

Sudipta Das

Assistant Professor,  
Department of Data Science,  
Ramakrishna Mission Vivekananda University, Kolkata  
Slides adapted from Jhonson & Winchern

## 1 Factor Analysis

# Introduction I

- Goal: To describe the covariance relationships among many variables in terms of a few underlying, but unobservable, random quantities called factors.
- Grouping of variables such that: intra-group correlation is high and inter-group correlation is low.
- Each group represents a single underlying factor, that is responsible for the observed correlations.

# Factor Model I

- Observable random vector:  $\mathbf{X}_{p \times 1}$ 
  - Mean vector:  $\mu_{p \times 1}$
  - Covariance Matrix:  $\Sigma_{p \times p}$
- Postulate:  $\mathbf{X}$  is linearly dependent upon
  - A few ( $m < p$ ) unobservable common factors  $F_1, \dots, F_m$
  - Additional  $p$  sources of specific variation  $\epsilon_1, \dots, \epsilon_p$
- Model

$$\mathbf{X} - \mu_{(p \times 1)} = L_{(p \times m)} \mathbf{F}_{(m \times 1)} + \epsilon_{(p \times 1)}$$

- $L$ : Loading matrix
  - $l_{ij}$ : loading of the  $i$ th variable on the  $j$ th factor
- Unobservable (random) variables are  $F_1, \dots, F_m, \epsilon_1, \dots, \epsilon_p$

# Orthogonal Factor Model: Assumptions I

- On Factors:  $E(\mathbf{F}) = \mathbf{0}_{m \times 1}$  and  $Cov(\mathbf{F}) = E[\mathbf{F}\mathbf{F}'] = \mathbf{I}_{m \times m}$
- On Errors:  $E(\epsilon) = \mathbf{0}_{p \times 1}$  and

$$Cov(\epsilon) = E[\epsilon\epsilon'] = \psi_{p \times p} = \begin{bmatrix} \psi_1 & 0 & \dots & 0 \\ 0 & \psi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \psi_p \end{bmatrix}$$

- On dependency:  $Cov(\epsilon, \mathbf{F}) = E[\epsilon\mathbf{F}'] = \mathbf{0}_{p \times m}$

# Orthogonal Factor Model: Outcomes I

- $Cov(\mathbf{X}) = \Sigma = LL' + \psi$
- $Cov(\mathbf{X}, \mathbf{F}) = L$
- $Var(X_i) = \sigma_{ii} = l_{i1}^2 + l_{i2}^2 + \dots + l_{im}^2 + \psi_i$ 
  - $Var(X_i)$  = communality+specific variance
- Example 9.1 (Page 484)

# Orthogonal Factor Model: Outcomes II

- Non-existence of proper solution

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

- Example 9.2 (Page 486)
- Solution is not unique:  $L^* = LT$

# Method of Estimation: Principal Component Analysis I

- Spectral decomposition:

$$\begin{aligned}\Sigma &= \lambda_1 \mathbf{e}_1 \mathbf{e}_1' + \lambda_2 \mathbf{e}_2 \mathbf{e}_2' + \dots + \lambda_p \mathbf{e}_p \mathbf{e}_p' \\ &= \begin{bmatrix} \sqrt{\lambda_1} \mathbf{e}_1 & \sqrt{\lambda_2} \mathbf{e}_2 & \dots & \sqrt{\lambda_p} \mathbf{e}_p \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} \mathbf{e}_1' \\ \sqrt{\lambda_2} \mathbf{e}_2' \\ \vdots \\ \sqrt{\lambda_p} \mathbf{e}_p' \end{bmatrix}\end{aligned}$$

- $\lambda_1 \geq \lambda_2 \geq \dots \lambda_p$
- It contains all factors i.e.,  $m = p$ 
  - $\Sigma = LL'$
  - $\psi_i = 0$  for  $i = 1, \dots, p$
- Exact Solution



# Method of Estimation: Principal Component Analysis II

- To identify a small number ( $m < p$ ) of factors
  - Ignore contributions of last small ( $p - m$ ) eigenvalues
- Steps:
  - Consider first  $m$  largest eigenvalues:

$$\begin{aligned}\Sigma &\approx \lambda_1 \mathbf{e}_1 \mathbf{e}_1' + \lambda_2 \mathbf{e}_2 \mathbf{e}_2' + \dots + \lambda_m \mathbf{e}_m \mathbf{e}_m' \\ &= \begin{bmatrix} \sqrt{\lambda_1} \mathbf{e}_1 & \sqrt{\lambda_2} \mathbf{e}_2 & \dots & \sqrt{\lambda_m} \mathbf{e}_m \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} \mathbf{e}_1' \\ \sqrt{\lambda_2} \mathbf{e}_2' \\ \vdots \\ \sqrt{\lambda_m} \mathbf{e}_m' \end{bmatrix}\end{aligned}$$

# Method of Estimation: Principal Component Analysis III

- Introduce specific factors:

$$\begin{aligned}\Sigma &\approx \begin{bmatrix} \sqrt{\lambda_1}\mathbf{e}_1 & \sqrt{\lambda_2}\mathbf{e}_2 & \cdots & \sqrt{\lambda_m}\mathbf{e}_m \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1}\mathbf{e}'_1 \\ \sqrt{\lambda_2}\mathbf{e}'_2 \\ \vdots \\ \sqrt{\lambda_m}\mathbf{e}'_m \end{bmatrix} + \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_p \end{bmatrix} \\ &= LL' + \Psi\end{aligned}$$

- Specific factor for  $i$ th feature:

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

# Method of Estimation: PCA Estimation I

- Sample covariance (correlation) matrix  $S(R)$
- Matrix of Estimated Factor Loading:

$$\tilde{L} = \left[ \sqrt{\hat{\lambda}_1} \hat{e}_1 : \sqrt{\hat{\lambda}_2} \hat{e}_2 : \dots : \sqrt{\hat{\lambda}_m} \hat{e}_m \right]$$

- $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \hat{\lambda}_m$
- Estimated specific variance:

$$\tilde{\psi}_i = s_{ii} - \sum_{k=1}^m \tilde{l}_{ik}^2 \text{ or } \left( \tilde{\psi}_i = 1 - \sum_{k=1}^m \tilde{l}_{ik}^2 \right)$$

- Residual matrix:

$$S - \tilde{L}\tilde{L}' - \tilde{\Psi} \text{ or } (R - \tilde{L}\tilde{L}' - \tilde{\Psi})$$

- Remarks:

- Sum of squared entries of  $(S(R) - \tilde{L}\tilde{L}' - \tilde{\Psi}) \leq \hat{\lambda}_{m+1}^2 + \dots + \hat{\lambda}_p^2$
- $\tilde{l}_{k1}^2 + \tilde{l}_{k2}^2 + \dots + \tilde{l}_{kp}^2 = \hat{\lambda}_k$
- Proportion of the total sample variance due to  $j$ th factor  
$$= \frac{\hat{\lambda}_j}{s_{11} + \dots + s_{pp}} \text{ or } \left( \frac{\hat{\lambda}_j}{p} \right)$$

- Example 9.3 (Page 491)

- Example 9.4 (Page 493)

# Method of Estimation: Maximum Likelihood Estimation

- Assumption: Normality
  - $F$  and  $\epsilon$  are normally distributed
- Likelihood:

$$L(\mu, \Sigma) = (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} e^{-\frac{1}{2} \text{tr} \left[ \Sigma^{-1} \left( \sum_{j=1}^n (x_j - \bar{x})(x_j - \bar{x})' + n(\bar{x} - \mu)(\bar{x} - \mu)' \right) \right]}$$

- Assumption: Uniqueness

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

- $\tilde{L}$  and  $\tilde{\Psi}$  are obtained by maximizing the likelihood numerically.

# Method of Estimation: Maximum Likelihood Estimation II

- To factorize  $R$ , replace  $X$  by  $Z = V^{-\frac{1}{2}}(X - \mu)$  in the likelihood, where

$$V = \begin{bmatrix} \sigma_{11} & 0 & \dots & 0 \\ 0 & \sigma_{22} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \sigma_{pp} \end{bmatrix}$$

- $\tilde{L}_Z$  and  $\tilde{\Psi}_Z$  are obtained by maximizing the likelihood numerically.
- Example 9.5 (Page 497)
- Example 9.6 (Page 499)

# Method of Estimation: Maximum Likelihood Estimation

## III

A Large sample test for the number of common factors

- We assume  $m$  common factor model holds (Null hypothesis)

$$H_0 : \Sigma_{p \times p} = L_{p \times m} L'_{m \times p} + \Psi_{p \times p}$$

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

- Assumption: Population is normal
- We perform *likelihood ratio test* to test  $H_0$

# Method of Estimation: Maximum Likelihood Estimation IV

- Test statistic

$$-2 \ln \Lambda = -2 \ln \left[ \frac{\text{m.l. under } H_0}{\text{m.l. over whole space}} \right]$$

- Note

- 1  $\text{m.l. over whole space} = \frac{1}{(2\pi)^{np/2}} |S_n|^{-n/2} e^{-np/2}$ , with d.f.

$$\nu = p + p(p+1)/2$$

- 2 and

$$\text{m.l. under } H_0 = \frac{1}{(2\pi)^{np/2}} |\hat{\Sigma}|^{-n/2} e^{\left( -\frac{1}{2} \text{tr} \left[ \hat{\Sigma}^{-1} \left( \sum_{j=1}^n (x_j - \bar{x})(x_j - \bar{x})' \right) \right] \right)},$$

$$\text{with d.f. } \nu_0 = p + [pm + p - \frac{1}{2}m(m-1)]$$



# Method of Estimation: Maximum Likelihood Estimation

## V

- Test statistic

$$\begin{aligned}-2 \ln \Lambda &= -2 \ln \left( \frac{|\hat{\Sigma}|}{|S_n|} \right)^{-n/2} + n[tr(\hat{\Sigma}^{-1} S_n) - p] \\ &= n \ln \left( \frac{|\hat{L}\hat{L}' + \hat{\Psi}|}{|S_n|} \right) + 0 \\ &\sim \chi^2_{\nu - \nu_0 = [(p-m)^2 - p - m]/2}\end{aligned}$$

# Method of Estimation: Maximum Likelihood Estimation VI

- Using Bartlett's correction, we reject  $H_0$  at the  $\alpha$  level of significance if

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

- Note:

- To apply the test, we need  $(p - m)^2 > p + m$
- The identity

$$\ln \frac{|\hat{L}\hat{L}' + \hat{\Psi}|}{|S_n|} = \ln \frac{|\hat{L}_Z\hat{L}_Z' + \hat{\Psi}_Z|}{|R|}$$

- Example 9.7 (Page 503)

- Factor rotation: Orthogonal transformation of factor loadings

$$L^* = LT$$

- Why: Original loading may not be readily interpretable
- Goal: Rotate to find a pattern of loadings such that each variable loads highly on a single factor and has small to moderate loadings on the remaining factors
- Example 9.8 (Page 505) (Figure 9.1)

# Factor Rotation II

- Varimax Criterion: Select the orthogonal transformation  $T$  that makes

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

as large as possible, where  $\tilde{l}_{ij}^* = \frac{\hat{l}_{ij}^*}{\hat{h}_i}$ .

- Interpretation:

- $V \propto \sum_{j=1}^m (\text{Variance of squares of scaled loading for } j\text{th factor})$

- Example 9.8 (Page 505) (Figure 9.1)
- Example 9.9 (Page 508) (Figure 9.2)
- Example 9.10 (Page 510)
- Example 9.11 (Page 511)

# Factor Score I

- Factor Score: Estimated values of common (unobservable) factors
- For  $j$ th,  $j = 1, \dots, n$ , observation the factor scores:

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

- To estimate  $\hat{\mathbf{f}}_j$ , we assume
  - estimated factor loadings  $\hat{l}_{ij}$  and specific variance  $\psi_i$ , as if they were the true values.
- Two methods to find factor scores:
  - Weighted least squares method
  - Regression method

- Weighted least squares method

- Model:

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

- Solution:

$$\hat{\mathbf{F}} = (\mathbf{L}'\Psi^{-1}\mathbf{L})^{-1}\mathbf{L}'\Psi^{-1}(\mathbf{X} - \mu) = \Delta^{-1}\mathbf{L}'\Psi^{-1}(\mathbf{X} - \mu)$$

- Estimated factor scores for  $j^{th}$  observation

$$\hat{\mathbf{f}}_j = \hat{\Delta}^{-1}\hat{\mathbf{L}}'\hat{\Psi}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}}), j = 1, \dots, n.$$

# Factor Score III

- Estimated factor scores for  $j^{th}$  observation, while working with correlation matrix

$$\hat{\mathbf{f}}_j = \hat{\Delta}_z^{-1} \hat{L}'_z \hat{\Psi}^{-1} \mathbf{z}_j, j = 1, \dots, n.$$

- $\mathbf{z}_j = D^{-\frac{1}{2}}(\mathbf{x}_j - \bar{\mathbf{x}})$ , where

$$D = \begin{bmatrix} s_{11} & 0 & \dots & 0 \\ 0 & s_{22} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & s_{pp} \end{bmatrix}$$

# Factor Score IV

- Regression method

- Model:

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

- Assumption:  $\mathbf{F}$  and  $\epsilon$  are jointly normal as well as the others as mentioned earlier

- Solution:

- 1  $(\mathbf{X} - \mu) \sim N_p(\mathbf{0}, \mathbf{L}\mathbf{L}' + \Psi)$

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

- 3  $\begin{bmatrix} \mathbf{X} - \mu \\ \mathbf{F} \end{bmatrix} \sim N_{p+m}(\mathbf{0}, \Sigma^*), \text{ where}$

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

- 4  $\mathbf{F}|\mathbf{x} \sim N_m$  with

$$\text{mean} = E[\mathbf{F}|\mathbf{x}] = \mathbf{L}'\Sigma^{-1}(\mathbf{x} - \mu)$$

and

$$\text{covariance} = \text{Cov}(\mathbf{F}|\mathbf{x}) = \mathbf{I} - \mathbf{L}'\Sigma^{-1}\mathbf{L}$$



- Thus,

$$\hat{\mathbf{F}} = \hat{\mathbf{L}}' \hat{\Sigma}^{-1} (\mathbf{X} - \bar{\mathbf{X}}) = \hat{\mathbf{L}}' (\hat{\mathbf{L}} \hat{\mathbf{L}}' + \hat{\Psi})^{-1} (\mathbf{X} - \bar{\mathbf{X}})$$

- Estimated factor scores for  $j^{th}$  observation

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

- Estimated factor scores for  $j^{th}$  observation, while working with correlation matrix

$$\hat{\mathbf{f}}_j = \hat{\mathbf{L}}_z' \mathbf{R}^{-1} \mathbf{z}_j, j = 1, \dots, n.$$

# Factor Score VI

- Using the following identity:

$$\hat{\mathbf{L}}'(\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\Psi})^{-1} = (\mathbf{I} + \hat{\mathbf{L}}'\hat{\Psi}^{-1}\hat{\mathbf{L}})^{-1}\hat{\mathbf{L}}'\hat{\Psi}^{-1}, \quad [\text{ex 9.6}]$$

one can show

$$\begin{aligned}\hat{\mathbf{f}}_j^R &= \hat{\mathbf{L}}'(\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\Psi})^{-1}(\mathbf{x}_j - \bar{\mathbf{x}}) \\&= (\mathbf{I} + \hat{\mathbf{L}}'\hat{\Psi}^{-1}\hat{\mathbf{L}})^{-1}\hat{\mathbf{L}}'\hat{\Psi}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}}) \\&= (\mathbf{I} + \hat{\mathbf{L}}'\hat{\Psi}^{-1}\hat{\mathbf{L}})^{-1}(\hat{\mathbf{L}}'\hat{\Psi}^{-1}\hat{\mathbf{L}})\hat{\mathbf{f}}_j^{LS} \\&= (\mathbf{I} + (\hat{\mathbf{L}}'\hat{\Psi}^{-1}\hat{\mathbf{L}})^{-1})^{-1}\hat{\mathbf{f}}_j^{LS} \quad [\text{since, } (A + B)^{-1} = (I + B^{-1}A)B^{-1}] \\&= (\mathbf{I} + \hat{\Delta}^{-1})^{-1}\hat{\mathbf{f}}_j^{LS}\end{aligned}$$

and

$$\hat{\mathbf{f}}_j^{LS} = (\mathbf{I} + \hat{\Delta}^{-1})\hat{\mathbf{f}}_j^R$$