ME731 - Métodos em Análise Multivariada
– Análise Fatorial II –

Prof. Carlos Trucíos
ctrucios@unicamp.br
ctruciosm.github.io

Instituto de Matemática, Estatística e Computação Científica,
Universidade Estadual de Campinas

Aula 14

Agenda I

Introdução

Introdução

O objetivo da análise fatorial (AF) é explicar as p variaveis no conjunto de dados utilizando k *fatores* ($k << p$).

Introdução

Definição

Seja $\mathbf{X} \sim (\mu, \Sigma)$ um vetor aleatório p-dimensional. O modelo k-fatorial de $\mathbf{X}$ é dado por

$$\mathbf{X} = \mu + \Lambda \mathbf{F} + \mathbf{u},$$

em que o vetor aleatório $\mathbf{F} \in \mathbb{R}^k$ é chamado de *fator comum*, $\Lambda_{p\times k}$ são as *cargas fatoriais* e o vetor aleatório $\mathbf{u} \in \mathbb{R}^p$ é chamado de *fator específico* ou *fator idiossincrático*.

- $\mathbf{F} \sim (0, \mathbf{I})$.
- $\mathbf{u} \sim (0, \mathbf{\Psi}), \mathbf{\Psi} = Diag\{\psi_{11}, \cdots, \psi_{pp}\}$.
- $\mathbb{C}ov(\mathbf{F}, \mathbf{u}) = 0$.

Introdução

Definição

Seja $\mathbf{X} \sim (\mu, \Sigma)$ um vetor aleatório p-dimensional. O modelo k-fatorial de $\mathbf{X}$ é dado por

$$\mathbf{X} = \mu + \Lambda\mathbf{F} + \mathbf{u},$$

em que o vetor aleatório $\mathbf{F} \in \mathbb{R}^k$ é chamado de *fator comum*, $\Lambda_{p\times k}$ são as *cargas fatoriais* e o vetor aleatório $\mathbf{u} \in \mathbb{R}^p$ é chamado de *fator específico* ou *fator idiossincrático*.

- $\mathbf{F} \sim (0, \mathbf{I})$.
- $\mathbf{u} \sim (0, \mathbf{\Psi}), \mathbf{\Psi} = Diag\{\psi_{11}, \cdots, \psi_{pp}\}$.
- $\mathbb{C}ov(\mathbf{F}, \mathbf{u}) = 0$.

Alguns preferem chamar o modelo acima de *Modelo Fatorial Ortogonal*, aqui utilizaremos apenas *Modelo Fatorial*. color{black}

Introdução

$$\mathbf{X} = \mu + \Lambda \mathbf{F} + \mathbf{u}$$

implica que,

$$X_j = \mu_j + \sum_{l=1}^{k} \lambda_{jl} F_l + u_j, \quad j = 1, \cdots, p.$$

Introdução

$$\mathbf{X} = \mu + \Lambda\mathbf{F} + \mathbf{u}$$

implica que,

$$X_j = \mu_j + \sum_{l=1}^{k} \lambda_{jl} F_l + u_j, \quad j = 1, \cdots, p.$$

Então,

$$\mathbb{V}(X_j) = \underbrace{\sum_{l=1}^{k} \lambda_{jl}^2}_{h_j^2} + \psi_{jj} \quad e \quad \Sigma = \Lambda\Lambda' + \boldsymbol{\Psi}.$$

O modelo fatorial explica a maior parte da variação de **X** através de apenas k fatores. Esses fatores explicam completamente a correlação de **X**.

Análise Fatorial vía Componentes Principais

Análise Fatorial vía Componentes Principais

Seja $\mathbf{X} \sim (\mu, \Sigma)$ e seja o modelo k-fatorial dado por

$$\mathbf{X} = \mu + \Lambda\mathbf{F} + \mathbf{u}.$$

Análise Fatorial vía Componentes Principais

Seja $\mathbf{X} \sim (\mu, \Sigma)$ e seja o modelo *k*-fatorial dado por

$$\mathbf{X} = \mu + \Lambda \mathbf{F} + \mathbf{u}.$$

sem perda de generalidade, assumiremos que $\mu = 0$.

Análise Fatorial vía Componentes Principais

Seja $\mathbf{X} \sim (\mu, \Sigma)$ e seja o modelo k-fatorial dado por

$$\mathbf{X} = \mu + \Lambda \mathbf{F} + \mathbf{u}.$$

sem perda de generalidade, assumiremos que $\mu = 0$. Utilizando ACP temos que as componentes principais são dadas por

$$\mathbf{Z} = \mathbf{P}'\mathbf{X},$$

em que $\mathbf{P}$ é obtido através de $\Sigma = \mathbf{PDP}'$.

Análise Fatorial vía Componentes Principais

Seja $\mathbf{X} \sim (\mu, \Sigma)$ e seja o modelo k-fatorial dado por

$$\mathbf{X} = \mu + \Lambda \mathbf{F} + \mathbf{u}.$$

sem perda de generalidade, assumiremos que $\mu = 0$. Utilizando ACP temos que as componentes principais são dadas por

$$\mathbf{Z} = \mathbf{P}'\mathbf{X},$$

em que $\mathbf{P}$ é obtido através de $\Sigma = \mathbf{PDP}'$.

Premultiplicado ambos os lados por $\mathbf{P}$, $\mathbf{PZ} = \underbrace{\mathbf{PP}'}_{\mathbf{I}}\mathbf{X} = \mathbf{X}$

Análise Fatorial vía Componentes Principais

Seja $\mathbf{X} \sim (\mu, \Sigma)$ e seja o modelo k-fatorial dado por

$$\mathbf{X} = \mu + \Lambda\mathbf{F} + \mathbf{u}.$$

sem perda de generalidade, assumiremos que $\mu = 0$. Utilizando ACP temos que as componentes principais são dadas por

$$\mathbf{Z} = \mathbf{P}'\mathbf{X},$$

em que $\mathbf{P}$ é obtido através de $\Sigma = \mathbf{PDP}'$.

Premultiplicado ambos os lados por $\mathbf{P}$, $\mathbf{PZ} = \underbrace{\mathbf{PP}'}_{\mathbf{I}}\mathbf{X} = \mathbf{X}$

$$\mathbf{X} = \mathbf{PZ} = \mathbf{PD}^{1/2}\mathbf{D}^{-1/2}\mathbf{Z} = \underbrace{\mathbf{PD}^{1/2}}_{\Lambda}\underbrace{\mathbf{D}^{-1/2}\mathbf{P}'\mathbf{X}}_{\mathbf{F}} \tag{1}$$

Análise Fatorial vía Componentes Principais

O resultado em (1), embora correto, não é muito útil na prática pois temos tantos fatores quanto variáveis.

Análise Fatorial vía Componentes Principais

O resultado em (1), embora correto, não é muito útil na prática pois temos tantos fatores quanto variáveis.

Se fizermos $\Lambda = [\Lambda_1 \quad \Lambda_2]$ e $\mathbf{F} = \begin{bmatrix} \mathbf{F}_1 \\ \mathbf{F}_2 \end{bmatrix}$, podemos reescrever (1) como:

Análise Fatorial vía Componentes Principais

O resultado em (1), embora correto, não é muito útil na prática pois temos tantos fatores quanto variáveis.

Se fizermos $\Lambda = [\Lambda_1 \quad \Lambda_2]$ e $\mathbf{F} = \begin{bmatrix} \mathbf{F}_1 \\ \mathbf{F}_2 \end{bmatrix}$, podemos reescrever (1) como:

$$\mathbf{X}_{p\times 1} = \Lambda_{p\times p}\mathbf{F}_{p\times 1} = [\Lambda_1 \quad \Lambda_2] \times \begin{bmatrix} \mathbf{F}_1 \\ \mathbf{F}_2 \end{bmatrix} = \underbrace{\Lambda_1}_{p\times k} \underbrace{\mathbf{F}_1}_{k\times 1} + \underbrace{\Lambda_2\mathbf{F}_2}_{\mathbf{u}} \tag{2}$$

Análise Fatorial vía Componentes Principais

O resultado em (1), embora correto, não é muito útil na prática pois temos tantos fatores quanto variáveis.

Se fizermos $\Lambda = [\Lambda_1 \quad \Lambda_2]$ e $\mathbf{F} = \begin{bmatrix} \mathbf{F}_1 \\ \mathbf{F}_2 \end{bmatrix}$, podemos reescrever (1) como:

$$\mathbf{X}_{p\times 1} = \Lambda_{p\times p}\mathbf{F}_{p\times 1} = [\Lambda_1 \quad \Lambda_2] \times \begin{bmatrix} \mathbf{F}_1 \\ \mathbf{F}_2 \end{bmatrix} = \underbrace{\Lambda_1}_{p\times k} \underbrace{\mathbf{F}_1}_{k\times 1} + \underbrace{\Lambda_2\mathbf{F}_2}_{\mathbf{u}} \qquad (2)$$

Mas quem são Λ_1 e $\mathbf{F}_1$ em termos de $\mathbf{X}$, $\mathbf{P}$ e $\mathbf{D}$?

Análise Fatorial vía Componentes Principais

Se fizermos $\mathbf{P} = [\mathbf{P}_1 \quad \mathbf{P}_2]$ e $\mathbf{D} = \begin{bmatrix} \mathbf{D}_1 & 0 \\ 0 & \mathbf{D}_2 \end{bmatrix}$, temos que:

Análise Fatorial vía Componentes Principais

Se fizermos $\mathbf{P} = [\mathbf{P}_1 \quad \mathbf{P}_2]$ e $\mathbf{D} = \begin{bmatrix} \mathbf{D}_1 & 0 \\ 0 & \mathbf{D}_2 \end{bmatrix}$, temos que:

- $\mathbf{Z} = \mathbf{P}'\mathbf{X} = \begin{bmatrix} \mathbf{P}_1' \\ \mathbf{P}_2' \end{bmatrix} \mathbf{X} = \begin{bmatrix} \mathbf{P}_1'\mathbf{X} \\ \mathbf{P}_2'\mathbf{X} \end{bmatrix}$
- $\mathbf{F} = \mathbf{D}^{-1/2}\mathbf{Z} = \mathbf{D} = \begin{bmatrix} \mathbf{D}_1^{-1/2} & 0 \\ 0 & \mathbf{D}_2^{-1/2} \end{bmatrix} \begin{bmatrix} \mathbf{P}_1'\mathbf{X} \\ \mathbf{P}_2'\mathbf{X} \end{bmatrix} = \begin{bmatrix} \mathbf{D}_1^{-1/2}\mathbf{P}_1'\mathbf{X} \\ \mathbf{D}_2^{-1/2}\mathbf{P}_2'\mathbf{X} \end{bmatrix}$
- $\Lambda = \mathbf{P}\mathbf{D}^{1/2} = [\mathbf{P}_1 \quad \mathbf{P}_2] \begin{bmatrix} \mathbf{D}_1^{1/2} & 0 \\ 0 & \mathbf{D}_2^{1/2} \end{bmatrix} = [\mathbf{P}_1\mathbf{D}_1^{1/2} \quad \mathbf{P}_2\mathbf{D}_2^{1/2}]$

Então,

$$\mathbf{F}_1 = \mathbf{D}_1^{-1/2}\mathbf{P}_1'\mathbf{X}, \quad \Lambda_1 = \mathbf{P}_1\mathbf{D}_1^{1/2} \quad e \quad \mathbf{u} = \mathbf{P}_2\mathbf{D}_2^{1/2}\mathbf{D}_2^{-1/2}\mathbf{P}_2'\mathbf{X} \qquad (3)$$

Análise Fatorial vía Componentes Principais

Na práticas temos uma matriz de dados

$$\mathbf{x} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$$

(n realizações de $\mathbf{X} \sim (\mu, \Sigma)$ com μ e Σ desconhecidos).

Análise Fatorial vía Componentes Principais

Na práticas temos uma matriz de dados

$$\mathbf{x} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$$

(n realizações de $\mathbf{X} \sim (\mu, \Sigma)$ com μ e Σ desconhecidos).

Da decomposição espectral temos $\mathbf{P}$ e $\mathbf{D}$ ($\mathbf{S} = \mathbf{PDP}$) e de (3) temos que

$$\mathbf{F}_1 = \mathbf{D}_1^{-1/2}\mathbf{P}_1'\mathbf{X}, \quad \Lambda_1 = \mathbf{P}_1\mathbf{D}_1^{1/2} \quad e \quad \mathbf{u} = \mathbf{P}_2\mathbf{D}_2^{1/2}\mathbf{D}_2^{-1/2}\mathbf{P}_2'\mathbf{X}$$

Análise Fatorial vía Componentes Principais

$$\mathbf{F}_1 = \mathbf{D}_1^{-1/2}\mathbf{P}_1'\mathbf{X}, \quad \Lambda_1 = \mathbf{P}_1\mathbf{D}_1^{1/2} \quad e \quad \mathbf{u} = \mathbf{P}_2\mathbf{D}_2^{1/2}\mathbf{D}_2^{-1/2}\mathbf{P}_2'\mathbf{X}$$

No caso da matriz de dados,

- $\hat{F}_i' = \mathbf{D}_1^{-1/2}\mathbf{P}_1'\mathbf{x}_i' \to \underbrace{\hat{F}_i}_{1\times p} = \underbrace{\mathbf{x}_i}_{1\times p}\mathbf{P}_1\mathbf{D}_1^{-1/2} \to \underbrace{\hat{\mathbf{F}}}_{n\times p} = \underbrace{\mathbf{x}}_{n\times p}\mathbf{P}_1\mathbf{D}_1^{-1/2}$
- $\Lambda = \mathbf{P}_1\mathbf{D}_1^{1/2}$
- $\mathbf{u}_i' = \mathbf{P}_2\mathbf{P}_2'\mathbf{x}_i' \to \underbrace{\mathbf{u}_i}_{1\times p} = \underbrace{\mathbf{x}_i}_{1\times p}\underbrace{\mathbf{P}_2}_{p\times(p-k)}\underbrace{\mathbf{P}_2'}_{(p-k)\times p} \to \mathbf{u} = \mathbf{X}\mathbf{P}_2\mathbf{P}_2'.$

Análise Fatorial vía Componentes Principais

Os passos para estimar o modelo k-fatorial através de ACP podem ser resumidor como:

Algoritmo:

1. Estimar Σ por $\mathbf{S}$.
2. Aplicar a decomposição espectral $\mathbf{S} = \mathbf{PDP}$
3. Fazer $\hat{\mathbf{F}} = \mathbf{x}\mathbf{P}_1\mathbf{D}_1^{-1/2}$, $\hat{\Lambda} = \mathbf{P}_1\mathbf{D}_1^{1/2}$ e $\hat{\mathbf{u}} = \mathbf{x}\mathbf{P}_2\mathbf{P}_2'$

$\hat{\mathbf{u}}$ no passo 3 pode também ser obtido como $\mathbf{x} - \hat{\mathbf{F}}\hat{\Lambda}'$

Análise Fatorial vía Componentes Principais

Os retornos semanais de 5 ações negociadas no NYSE estão disponíveis aqui.

```
library(dplyr)
library(expm)
retornos <- read.table("https://raw.githubusercontent.com/ctruci
glimpse(retornos)
```

```
## Rows: 103
## Columns: 5
## $ V1 <dbl> 1.30338, 0.84862, -1.79153, 2.15589, 1.08225, 1.01
## $ V2 <dbl> -0.78431, 1.66886, -0.86393, -0.34858, 0.37167, -1
## $ V3 <dbl> -0.31889, -0.62100, 1.00360, 1.74353, -1.01345, -0
## $ V4 <dbl> -4.47693, 1.19560, 0.00000, -2.85917, 2.91900, 1.3
## $ V5 <dbl> 0.52151, 1.34890, -0.61428, -0.69534, 4.09751, 0.2
```

Análise Fatorial vía Componentes Principais

```
retornos <- scale(retornos, center = TRUE, scale = FALSE)
# Decomposição Espectral
eigen_aux <- eigen(cov(retornos))
P <- eigen_aux$vectors        # autovetores
D <- diag(eigen_aux$values)   # autovalores
```

Análise Fatorial vía Componentes Principais

```r
retornos <- scale(retornos, center = TRUE, scale = FALSE)
# Decomposição Espectral
eigen_aux <- eigen(cov(retornos))
P <- eigen_aux$vectors          # autovetores
D <- diag(eigen_aux$values)    # autovalores
```

Trabalharemos com $k = 2$ fatores.

```r
F_hat <- retornos %*% P[,1:2] %*% solve(sqrtm(D[1:2, 1:2]))
Lambda_hat <- P[, 1:2] %*% sqrtm(D[1:2, 1:2])
u_hat <- retornos - F_hat %*% t(Lambda_hat)
```

Análise Fatorial vía Componentes Principais

```
round(cov(F_hat), 4)
```

```
##      [,1] [,2]
## [1,]    1    0
## [2,]    0    1
```

```
round(cov(F_hat, u_hat), 4)
```

```
##      V1 V2 V3 V4 V5
## [1,]  0  0  0  0  0
## [2,]  0  0  0  0  0
```

Análise Fatorial vía Componentes Principais

```
round(cov(u_hat), 4)
```

```
##          V1      V2      V3      V4      V5
## V1  0.9128 -0.6803 -0.3918 -0.2191  0.3170
## V2 -0.6803  0.8145 -0.2287  0.1207 -0.2158
## V3 -0.3918 -0.2287  1.0798 -0.0198  0.0047
## V4 -0.2191  0.1207 -0.0198  1.2099 -1.1650
## V5  0.3170 -0.2158  0.0047 -1.1650  1.1359
```

$\mathbb{V}(\hat{\mathbf{u}})$ não é uma matriz diagional! "Tá" certo isso?

Análise Fatorial vía Componentes Principais

```
round(cov(u_hat), 4)
```

```
##          V1      V2      V3      V4      V5
## V1  0.9128 -0.6803 -0.3918 -0.2191  0.3170
## V2 -0.6803  0.8145 -0.2287  0.1207 -0.2158
## V3 -0.3918 -0.2287  1.0798 -0.0198  0.0047
## V4 -0.2191  0.1207 -0.0198  1.2099 -1.1650
## V5  0.3170 -0.2158  0.0047 -1.1650  1.1359
```

$\mathbb{V}(\hat{\mathbf{u}})$ não é uma matriz diagional! "Tá" certo isso? **Sim!** nada garante que $\mathbb{V}(\hat{\mathbf{u}})$ será diagional.

Análise Fatorial vía Componentes Principais

```r
round(cov(u_hat), 4)
```

```
##          V1      V2      V3      V4      V5
## V1   0.9128 -0.6803 -0.3918 -0.2191  0.3170
## V2  -0.6803  0.8145 -0.2287  0.1207 -0.2158
## V3  -0.3918 -0.2287  1.0798 -0.0198  0.0047
## V4  -0.2191  0.1207 -0.0198  1.2099 -1.1650
## V5   0.3170 -0.2158  0.0047 -1.1650  1.1359
```

$\mathbb{V}(\hat{\mathbf{u}})$ não é uma matriz diagonal! "Tá" certo isso? **Sim!** nada garante que $\mathbb{V}(\hat{\mathbf{u}})$ será diagonal.

Lembre-se, no AF aproximamos $\mathbf{S} \approx \hat{\Lambda}\hat{\Lambda}' + \mathbf{\Psi}$ (ou seja, estamos desconsiderando os elemtos de $\mathbb{V}(\hat{\mathbf{u}})$ que estão fora da diagonal).

Análise Fatorial vía Componentes Principais

Como determinar o número de fatores?

- Geralmente, ele é predefinido pelo pesquisador da área.
- Se ele não for previamente definido, podem ser utilizados os mesmos critérios utilizados em ACP.

Análise Fatorial vía Componentes Principais

Como determinar o número de fatores?

- Geralmente, ele é predefinido pelo pesquisador da área.
- Se ele não for previamente definido, podem ser utilizados os mesmos critérios utilizados em ACP.

Para avaliar se nossa escolha de k foi boa, olhamos para:

- h_i^2 (esperamos que a cumunalidade seja alta).
- $\mathbf{S} - \hat{\Lambda}_1\hat{\Lambda}_1'$ (esperamos que os elementos fora da diagonal sejam pequenos)

Máxima Versossimilhança

Máxima Versossimilhança

Seja $\mathbf{X} \sim N_p(\mu, \Sigma)$, então $\hat{\Lambda}$ e $\hat{\mathbf{\Psi}}$ podem ser obtidos pelo método de máxima verossimilhança.

Máxima Versossimilhança

Seja $\mathbf{X} \sim N_p(\mu, \Sigma)$, então $\hat{\Lambda}$ e $\hat{\mathbf{\Psi}}$ podem ser obtidos pelo método de máxima verossimilhança.

Como $\hat{\mu} = \bar{\mathbf{X}}$, a log-verosimilhança é reducida a,

$$l = -\frac{1}{2} n \log |2\pi\Sigma| - \frac{1}{2} n Tr(\Sigma^{-1}\mathbf{S}).$$

Máxima Versossimilhança

Seja $\mathbf{X} \sim N_p(\mu, \Sigma)$, então $\hat{\Lambda}$ e $\hat{\mathbf{\Psi}}$ podem ser obtidos pelo método de máxima verossimilhança.

Como $\hat{\mu} = \bar{\mathbf{X}}$, a log-verosimilhança é reducida a,

$$l = -\frac{1}{2} n \log |2\pi\Sigma| - \frac{1}{2} n Tr(\Sigma^{-1}\mathbf{S}).$$

Mas $\Sigma = \Lambda\Lambda' + \mathbf{\Psi}$, então podemos re-escrever a verosimilhança como

$$\begin{aligned} l &= -\frac{p}{2}\log(2\pi) - \frac{1}{2} n \log |\Lambda\Lambda' + \mathbf{\Psi}| - \frac{1}{2} n Tr([\Lambda\Lambda' + \mathbf{\Psi}]^{-1}\mathbf{S}) && (4) \\ &\propto -\frac{1}{2} n \log |\Lambda\Lambda' + \mathbf{\Psi}| - \frac{1}{2} n Tr([\Lambda\Lambda' + \mathbf{\Psi}]^{-1}\mathbf{S}) && (5) \\ &\propto -\log |\Lambda\Lambda' + \mathbf{\Psi}| - Tr([\Lambda\Lambda' + \mathbf{\Psi}]^{-1}\mathbf{S}) && (6) \end{aligned}$$

Máxima Versossimilhança

Devido à multiplicidade de $\Lambda\Lambda'$ a função a ser otimizada ainda não está bem definida. Para contornar esse problema, impomos a restrição de que $\Lambda\boldsymbol{\Psi}\Lambda'$ seja diagonal.

Máxima Versossimilhança

Devido à multiplicidade de $\Lambda\Lambda'$ a função a ser otimizada ainda não está bem definida. Para contornar esse problema, impomos a restrição de que $\Lambda\boldsymbol{\Psi}\Lambda'$ seja diagonal.

Assim, basta maximizar (3) [ou minimizar -(3)] *w.r.t.* Λ e $\boldsymbol{\Psi}$ e com a restrição imposta para obter os EMV.

Máxima Versossimilhança

Devido à multiplicidade de $\Lambda\Lambda'$ a função a ser otimizada ainda não está bem definida. Para contornar esse problema, impomos a restrição de que $\Lambda\boldsymbol{\Psi}\Lambda'$ seja diagonal.

Assim, basta maximizar (3) [ou minimizar -(3)] *w.r.t.* Λ e $\boldsymbol{\Psi}$ e com a restrição imposta para obter os EMV.

A solução é numérica

O problema? mesmo para $k = 1$ a maximização/minimização é bastante complexa e algoritmos iterativos são utilizados.

Fatores Estimados

Embora o foco principal esteja em estimar Λ e $\boldsymbol{\Psi}$, às vezes é de interesse obter $\hat{\mathbf{F}}$.

Fatores Estimados

Embora o foco principal esteja em estimar Λ e $\mathbf{\Psi}$, às vezes é de interesse obter $\hat{\mathbf{F}}$.

Quando utilizamos AF vía ACP, $\hat{\mathbf{F}} = \hat{\mathbf{D}}_1^{-1/2}\hat{\mathbf{P}}_1\mathbf{X}$. Em geral, existem outros métodos que podemos utilizar:

Fatores Estimados

Embora o foco principal esteja em estimar Λ e $\mathbf{\Psi}$, às vezes é de interesse obter $\hat{\mathbf{F}}$.

Quando utilizamos AF vía ACP, $\hat{\mathbf{F}} = \hat{\mathbf{D}}_1^{-1/2}\hat{\mathbf{P}}_1\mathbf{X}$. Em geral, existem outros métodos que podemos utilizar:

- Método de mínimo quadrados ponderados: $\hat{\mathbf{F}} = (\hat{\Lambda}'\hat{\mathbf{\Psi}}^{-1}\hat{\Lambda})^{-1}\hat{\Lambda}'\hat{\mathbf{\Psi}}^{-1}\mathbf{X}$
- Regressão: $\hat{\mathbf{F}} = \hat{\Lambda}'\mathbf{S}^{-1}\mathbf{X}$
- etc.

Rotação

Rotação

- As restrições impostas na estimação ($\Lambda'\boldsymbol{\Psi}^{-1}\Lambda$ ou $\Lambda'\mathbf{D}^{-1}\Lambda$ ser diagonal), são apenas formas matemáticas de lidar com o problema de multiplicidade.

Rotação

- As restrições impostas na estimação ($\Lambda'\mathbf{\Psi}^{-1}\Lambda$ ou $\Lambda'\mathbf{D}^{-1}\Lambda$ ser diagonal), são apenas formas matemáticas de lidar com o problema de multiplicidade.
- Contudo, essas restrições não garantem que a interpretação dos fatores seja fácil.

Rotação

- As restrições impostas na estimação ($\Lambda'\mathbf{\Psi}^{-1}\Lambda$ ou $\Lambda'\mathbf{D}^{-1}\Lambda$ ser diagonal), são apenas formas matemáticas de lidar com o problema de multiplicidade.
- Contudo, essas restrições não garantem que a interpretação dos fatores seja fácil.
- Utilizando a não unicidade da solução podemos encontrar matrizes ortogonais que nos levarão a uma melhor interpretação.

Rotação

- As restrições impostas na estimação ($\Lambda'\mathbf{\Psi}^{-1}\Lambda$ ou $\Lambda'\mathbf{D}^{-1}\Lambda$ ser diagonal), são apenas formas matemáticas de lidar com o problema de multiplicidade.
- Contudo, essas restrições não garantem que a interpretação dos fatores seja fácil.
- Utilizando a não unicidade da solução podemos encontrar matrizes ortogonais que nos levarão a uma melhor interpretação.
- Utilizar matrizes ortogonais significa *rotar os fatores* de forma que sejam mais facilmente interpretáveis

Rotação

- As restrições impostas na estimação ($\Lambda'\boldsymbol{\Psi}^{-1}\Lambda$ ou $\Lambda'\mathbf{D}^{-1}\Lambda$ ser diagonal), são apenas formas matemáticas de lidar com o problema de multiplicidade.
- Contudo, essas restrições não garantem que a interpretação dos fatores seja fácil.
- Utilizando a não unicidade da solução podemos encontrar matrizes ortogonais que nos levarão a uma melhor interpretação.
- Utilizar matrizes ortogonais significa *rotar os fatores* de forma que sejam mais facilmente interpretáveis
- Existem vários métodos para fazer a rotação.

Rotação

- Seja $\hat{F}$ a solução inicial (obtida vía ACP, MV ou algúm outro método).

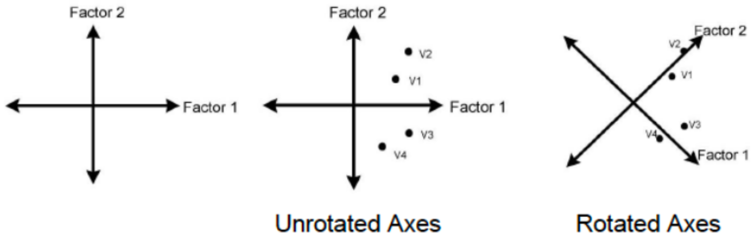
Rotação

- Seja $\hat{F}$ a solução inicial (obtida vía ACP, MV ou algúm outro método).
- Sabemos que $\mathbb{C}or(\hat{F}, \mathbf{X}) = Diag(\mathbf{S})^{-1/2}\hat{\Lambda}$ (ou apenas $\hat{\Lambda}$ se $\mathbf{X}$ for padronizada)

Rotação

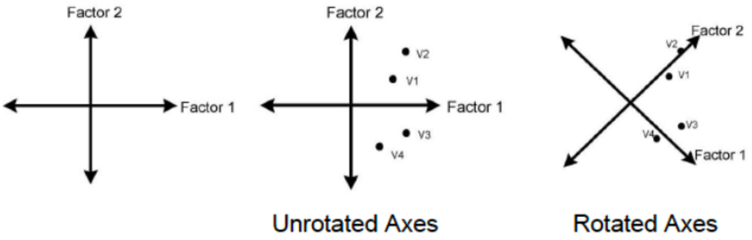
- Seja $\hat{F}$ a solução inicial (obtida vía ACP, MV ou algúm outro método).
- Sabemos que $\mathbb{C}or(\hat{F}, \mathbf{X}) = Diag(\mathbf{S})^{-1/2}\hat{\Lambda}$ (ou apenas $\hat{\Lambda}$ se $\mathbf{X}$ for padronizada)
- A ideia da rotação é que as variáveis originais tenham uma correlação o mais próximo de 1 com um fator e o mais próximo de 0 com os outros fatores (assim, cada fator terá um grupo de variáveis fortemente correlacionadas com ele).

Rotação



Fonte: https://clemson.instructure.com/files/2156796/download?download_frd=1

Rotação



Fonte: https://clemson.instructure.com/files/2156796/download?download_frd=1

A seguir, veremos o método **varimax** (um dos métodos de rotação mais conhecidos).

Rotação: Varimax

Pense em $k = 2$, a matriz de rotação (no sentido horário) é dada por

$$G(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

Rotação: Varimax

Pense em $k = 2$, a matriz de rotação (no sentido horário) é dada por

$$G(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

A matriz de cargas (já rotada) é dada por $\hat{\Lambda}^* = \hat{\Lambda}G(\theta)$.

Rotação: Varimax

Pense em $k = 2$, a matriz de rotação (no sentido horário) é dada por

$$G(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

A matriz de cargas (já rotada) é dada por $\hat{\Lambda}^* = \hat{\Lambda}G(\theta)$.

A ideia de **varimax** é encontrar θ de forma que maximize a soma das variâncias dos quadrados das cargas fatoriais (padronizadas pela cumunalidade) dentro de cada coluna de $\hat{\Lambda}^*$.

Rotação: Varimax

A variância dos quadrados das cargas fatoriais (padronizadas pela cumunalidade) do j-ésimo fator (rotado) é dada por,

$$\frac{\sum_{i=1}^{p}(\lambda_{ij}^{2*}/h_j^{2*})^2}{p} - \left[\frac{\sum_{i=1}^{p}\lambda_{ij}^{2*}/h_j^{2*}}{p}\right]^2$$

Rotação: Varimax

A variância dos quadrados das cargas fatoriais (padronizadas pela cumunalidade) do j-ésimo fator (rotado) é dada por,

$$\frac{\sum_{i=1}^{p}(\lambda_{ij}^{2*}/h_j^{2*})^2}{p} - \left[\frac{\sum_{i=1}^{p}\lambda_{ij}^{2*}/h_j^{2*}}{p}\right]^2$$

Varimax: escolhemos θ de forma que

$$\frac{1}{p}\sum_{j=1}^{k}\left[\sum_{i=1}^{p}\left(\frac{\lambda_{ij}^{2*}}{h_i^{2*}}\right)^2 - \frac{1}{p}\left(\sum_{i=1}^{p}\frac{\lambda_{ij}^{2*}}{h_i^{2*}}\right)^2\right] \quad \text{seja máximo.}$$

Rotação: Varimax

A variância dos quadrados das cargas fatoriais (padronizadas pela cumunalidade) do j-ésimo fator (rotado) é dada por,

$$\frac{\sum_{i=1}^{p}(\lambda_{ij}^{2*}/h_j^{2*})^2}{p} - \left[\frac{\sum_{i=1}^{p}\lambda_{ij}^{2*}/h_j^{2*}}{p}\right]^2$$

Varimax: escolhemos θ de forma que

$$\frac{1}{p}\sum_{j=1}^{k}\left[\sum_{i=1}^{p}\left(\frac{\lambda_{ij}^{2*}}{h_i^{2*}}\right)^2 - \frac{1}{p}\left(\sum_{i=1}^{p}\frac{\lambda_{ij}^{2*}}{h_i^{2*}}\right)^2\right] \quad \text{seja máximo.}$$

Observação: alguns chamam de **varimax padronizado**.

Rotação

Existem vários métodos de rotação:

- **Varimax:** (simplifica a interpretação dos fatores).
- **Quartimax:** minimiza o número de fatores necessários para explicar cada variável (simplifica a interpretação das variáveis observadas).
- **Equamax:** combinação entre varimax e quartimax.
- **Oblimin:** rotação obliqua (não ortogonal)
- **Promax:** rotação obliqua (mais rápida que oblimin)

Referências

Referências

- Härdle, W. K., & Simar, L. (2019). Applied Multivariate Statistical Analysis. Fifth Editon. Springer Nature. Capítulo 12.
- Mardia, K. V., Kent, J. T., & Bibby, J, M. (1979). Multivariate Analysis. Academic Press. Capítulo 9.
- Peña, D. (2002). Análisis de Datos Multivariantes. Mc Graw Hill. Capítulo 12.