SME0822 Análise Multivariada e Aprendizado Não-Supervisionado

Aula 8a: **Análise Fatorial**

Prof. Cibele Russo

cibele@icmc.usp.br

http://www.icmc.usp.br/~cibele

Johnson, R. A., & Wichern, D. W. (2007). Applied Multivariate Statistical Analysis. Prentice Hall.

Mingoti, S. A. (2007). Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada. Editora UFMG.

Análise fatorial

Objetivos:

- Descrever a variabilidade original do vetor aleatório $\underset{\sim}{X}_{p\times 1}$ em termos de um número menor m de variáveis aleatórias chamadas **fatores comuns**, e que estão relacionadas com $\underset{\sim}{X}$ por meio de um **modelo linear**.
- Agrupar as variáveis originais $X_i,\ i = 1, \ldots, p$, em fatores mutuamente não correlacionados, interpretáveis.
- Reduzir a dimensionalidade do problema.

Análise fatorial

Objetivos:

- Descrever a variabilidade original do vetor aleatório $\underset{\sim}{X}_{p\times 1}$ em termos de um número menor m de variáveis aleatórias chamadas **fatores comuns**, e que estão relacionadas com $\underset{\sim}{X}$ por meio de um **modelo linear**.
- Agrupar as variáveis originais $X_i,\ i = 1, \ldots, p$, em fatores mutuamente não correlacionados, interpretáveis.
- Reduzir a dimensionalidade do problema.

Análise fatorial

Objetivos:

- Descrever a variabilidade original do vetor aleatório $\underset{\sim}{X}_{p\times 1}$ em termos de um número menor m de variáveis aleatórias chamadas **fatores comuns**, e que estão relacionadas com $\underset{\sim}{X}$ por meio de um **modelo linear**.
- Agrupar as variáveis originais $X_i,\ i = 1, \ldots, p$, em fatores mutuamente não correlacionados, interpretáveis.
- Reduzir a dimensionalidade do problema.

Análise fatorial

Origem: Spearman (1904), em tentativas de **definir e medir inteligência**.

Áreas de aplicação originais: psicologia, psicometria, ciências sociais.

Modelo de análise fatorial via matriz de correlação

Seja $\underset{\sim}{X}_p$ um vetor aleatório com

$$\mathsf{E}(\underset{\sim}{X}) = \underset{\sim}{\mu},\ \mathsf{Var}(\underset{\sim}{X}) = \Sigma \text{ e } \mathsf{Cor}(\underset{\sim}{X}) = P.$$

Sejam

$$Z_i = \frac{X_i - \mu_i}{\sigma_i}, i = 1, \ldots, p$$

as variáveis padronizadas.

Já vimos que $\mathsf{Cov}(\underset{\sim}{Z}) = P = \mathsf{Cor}(\underset{\sim}{X})$.

Modelo de análise fatorial via matriz de correlação

Seja $\underset{\sim}{X}_p$ um vetor aleatório com

$$\mathsf{E}(\underset{\sim}{X}) = \underset{\sim}{\mu},\ \mathsf{Var}(\underset{\sim}{X}) = \Sigma \text{ e } \mathsf{Cor}(\underset{\sim}{X}) = P.$$

Sejam

$$Z_i = \frac{X_i - \mu_i}{\sigma_i}, i = 1, \ldots, p$$

as variáveis padronizadas.

Já vimos que $\mathsf{Cov}(\underset{\sim}{Z}) = P = \mathsf{Cor}(\underset{\sim}{X})$.

Modelo de análise fatorial via matriz de correlação

Seja $\underset{\sim}{X}_p$ um vetor aleatório com

$$\mathsf{E}(\underset{\sim}{X}) = \underset{\sim}{\mu},\ \mathsf{Var}(\underset{\sim}{X}) = \Sigma \text{ e } \mathsf{Cor}(\underset{\sim}{X}) = P.$$

Sejam

$$Z_i = \frac{X_i - \mu_i}{\sigma_i}, i = 1, \dots, p$$

as variáveis padronizadas.

Já vimos que $\mathsf{Cov}(\underset{\sim}{Z}) = P = \mathsf{Cor}(\underset{\sim}{X})$.

Modelo de análise fatorial via matriz de correlação

Queremos explicar as variáveis padronizadas $Z_1, \ldots, Z_p$ com m fatores $F_1, \ldots, F_m$ de tal forma que

$$Z_1 = l_{11}F_1 + \ldots + l_{1m}F_m + \epsilon_1$$

$$Z_2 = l_{21}F_1 + \ldots + l_{2m}F_m + \epsilon_2$$

$$\vdots$$

$$Z_p = l_{p1}F_1 + \ldots + l_{pm}F_m + \epsilon_p$$

Modelo de análise fatorial via matriz de correlação

Em notação matricial, podemos escrever

$$\underset{\sim}{Z} = L\underset{\sim}{F} + \underset{\sim}{\epsilon},$$

em que os elementos são dados por

$$\begin{pmatrix} Z_1 \\ \vdots \\ Z_p \end{pmatrix} = \begin{pmatrix} l_{11} & \dots & l_{1m} \\ \vdots & \ddots & \vdots \\ l_{p1} & \dots & l_{pm} \end{pmatrix} \begin{pmatrix} F_1 \\ \vdots \\ F_m \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_p \end{pmatrix}.$$

Modelo de análise fatorial via matriz de correlação

Nesse modelo,

- $\underset{\sim}{F}_{m\times 1}$ é um vetor aleatório que contém m fatores, também chamados de **variáveis latentes**, que descrevem os elementos da população e são não observáveis.
- $\underset{\sim}{\epsilon}$ é o vetor de erros aleatórios não observáveis.
- l_{ij}, também chamado de *loading* (peso) ou carga fatorial, é o coeficiente da i-ésima variável padronizada Z_i, $i = 1, \ldots, p$ no j-ésimo fator, F_j, $j = 1, \ldots, m$.
- L é a matriz de cargas fatoriais

Modelo de análise fatorial via matriz de correlação

Nesse modelo,

- $\underset{\sim}{F}_{m\times 1}$ é um vetor aleatório que contém m fatores, também chamados de **variáveis latentes**, que descrevem os elementos da população e são não observáveis.
- $\underset{\sim}{\epsilon}$ é o vetor de erros aleatórios não observáveis.
- l_{ij}, também chamado de *loading* (peso) ou carga fatorial, é o coeficiente da i-ésima variável padronizada Z_i, $i = 1, \ldots, p$ no j-ésimo fator, F_j, $j = 1, \ldots, m$.
- L é a matriz de cargas fatoriais

Modelo de análise fatorial via matriz de correlação

Nesse modelo,

- $\underset{\sim}{F}_{m\times 1}$ é um vetor aleatório que contém m fatores, também chamados de **variáveis latentes**, que descrevem os elementos da população e são não observáveis.
- $\underset{\sim}{\epsilon}$ é o vetor de erros aleatórios não observáveis.
- l_{ij}, também chamado de *loading* (peso) ou carga fatorial, é o coeficiente da i-ésima variável padronizada Z_i, $i = 1, \ldots, p$ no j-ésimo fator, F_j, $j = 1, \ldots, m$.
- L é a matriz de cargas fatoriais

Modelo de análise fatorial via matriz de correlação

Nesse modelo,

- $\underset{\sim}{F}_{m\times 1}$ é um vetor aleatório que contém m fatores, também chamados de **variáveis latentes**, que descrevem os elementos da população e são não observáveis.
- $\underset{\sim}{\epsilon}$ é o vetor de erros aleatórios não observáveis.
- l_{ij}, também chamado de *loading* (peso) ou carga fatorial, é o coeficiente da i-ésima variável padronizada Z_i, $i = 1, \ldots, p$ no j-ésimo fator, F_j, $j = 1, \ldots, m$.
- L é a matriz de cargas fatoriais

Modelo fatorial ortogonal

O modelo

$$\underset{\sim}{Z} = L\underset{\sim}{F} + \underset{\sim}{\epsilon},$$

com as suposições

i. $\text{E}(\underset{\sim}{F}) = \underset{\sim}{0}$

ii. $\text{Var}(\underset{\sim}{F}) = I$

iii. $\text{E}(\underset{\sim}{\epsilon}) = \underset{\sim}{0}$

iv. $\text{Var}(\underset{\sim}{\epsilon}) = \Psi = diag(\psi_1, \ldots, \psi_p)$

é chamado de **modelo fatorial ortogonal**, já que os m fatores são ortogonais entre si. Além disso, supomos que $\underset{\sim}{F}$ e $\underset{\sim}{\epsilon}$ são não correlacionados.

Modelo fatorial ortogonal

Resultado

Em um modelo fatorial ortogonal (suposições i a iv satisfeitas), pode-se escrever a matriz P como

$$P = LL^\top + \Psi.$$

Prova: Temos que

$$\begin{aligned} P &= \text{Var}(\underset{\sim}{Z}) \\ &= \text{Var}(L\underset{\sim}{F} + \underset{\sim}{\epsilon}) \\ &= L\text{Var}\underset{\sim}{F}L^\top + \text{Var}(\underset{\sim}{\epsilon}) = \\ &= LIL^\top + \Psi = LL^\top + \Psi. \end{aligned}$$

Modelo fatorial ortogonal

Resultado

Em um modelo fatorial ortogonal (suposições i a iv satisfeitas), pode-se escrever a matriz P como

$$P = LL^\top + \Psi.$$

Prova: Temos que

$$\begin{aligned} P &= \text{Var}(\underset{\sim}{Z}) \\ &= \text{Var}(L\underset{\sim}{F} + \underset{\sim}{\epsilon}) \\ &= L\text{Var}\underset{\sim}{F}L^\top + \text{Var}(\underset{\sim}{\epsilon}) = \\ &= LIL^\top + \Psi = LL^\top + \Psi. \end{aligned}$$

Modelo fatorial ortogonal

Objetivo:

Queremos encontrar $L_{p\times m}$ e $\Psi_{p\times p}$ que possam decompor a matriz P na forma

$$P = LL^\top + \Psi,$$

o que nem sempre é possível.

Modelo fatorial ortogonal

Note que

$$P = LL^{\top} + \Psi =$$

$$\begin{pmatrix} \sum_{j=1}^{m} l_{1j}^2 & \dots & \sum_{j=1}^{m} l_{1j} l_{jp} \\ \vdots & \ddots & \vdots \\ \sum_{j=1}^{m} l_{pj} l_{j1} & \dots & \sum_{j=1}^{m} l_{pj}^2 \end{pmatrix} + \begin{pmatrix} \psi_1 & 0 & \dots & 0 \\ 0 & \psi_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \psi_p \end{pmatrix}$$

Modelo fatorial ortogonal

Assim, vem que

- Var$(Z_i) = \sum_{j=1}^{m} l_{ij}^2 + \psi_i = h_i^2 + \psi_i = 1$, em que h_i^2 é chamado **comunalidade** (devido aos fatores) e ψ_i é chamado de **variância específica**.
- Cov$(Z_i, Z_k) = l_{i1}l_{1k} + \ldots + l_{im}l_{mk}$ para $i, k = 1, \ldots, p$ e $i \neq k$.
- Cov$(\underset{\sim}{Z}, \underset{\sim}{F}) = L$ e Cov$(Z_i, F_k) =$ Cor$(Z_i, F_k) = l_{ik}$ pois Cov$(\underset{\sim}{Z}, \underset{\sim}{F}) =$ Cov$(L\underset{\sim}{F} + \underset{\sim}{\epsilon}, \underset{\sim}{F}) =$ Cov$(L\underset{\sim}{F}, \underset{\sim}{F}) +$ Cov$(\underset{\sim}{\epsilon}, \underset{\sim}{F}) = L$.

Modelo fatorial ortogonal

Assim, vem que

- $\text{Var}(Z_i) = \sum_{j=1}^{m} l_{ij}^2 + \psi_i = h_i^2 + \psi_i = 1$, em que h_i^2 é chamado **comunalidade** (devido aos fatores) e ψ_i é chamado de **variância específica**.
- $\text{Cov}(Z_i, Z_k) = l_{i1}l_{1k} + \ldots + l_{im}l_{mk}$ para $i, k = 1, \ldots, p$ e $i \neq k$.
- $\text{Cov}(\underset{\sim}{Z}, \underset{\sim}{F}) = L$ e $\text{Cov}(Z_i, F_k) = \text{Cor}(Z_i, F_k) = l_{ik}$ pois $\text{Cov}(\underset{\sim}{Z}, \underset{\sim}{F}) = \text{Cov}(L\underset{\sim}{F} + \underset{\sim}{\epsilon}, \underset{\sim}{F}) = \text{Cov}(L\underset{\sim}{F}, \underset{\sim}{F}) + \text{Cov}(\underset{\sim}{\epsilon}, \underset{\sim}{F}) = L$.

Modelo fatorial ortogonal

Assim, vem que

- $\text{Var}(Z_i) = \sum_{j=1}^{m} l_{ij}^2 + \psi_i = h_i^2 + \psi_i = 1$, em que h_i^2 é chamado **comunalidade** (devido aos fatores) e ψ_i é chamado de **variância específica**.
- $\text{Cov}(Z_i, Z_k) = l_{i1}l_{1k} + \ldots + l_{im}l_{mk}$ para $i, k = 1, \ldots, p$ e $i \neq k$.
- $\text{Cov}(\underset{\sim}{Z}, \underset{\sim}{F}) = L$ e $\text{Cov}(Z_i, F_k) = \text{Cor}(Z_i, F_k) = l_{ik}$ pois $\text{Cov}(\underset{\sim}{Z}, \underset{\sim}{F}) = \text{Cov}(L\underset{\sim}{F} + \underset{\sim}{\epsilon}, \underset{\sim}{F}) = \text{Cov}(L\underset{\sim}{F}, \underset{\sim}{F}) + \text{Cov}(\underset{\sim}{\epsilon}, \underset{\sim}{F}) = L$.

Modelo fatorial ortogonal

Assim, vem que

- $\text{Var}(Z_i) = \sum_{j=1}^{m} l_{ij}^2 + \psi_i = h_i^2 + \psi_i = 1$, em que h_i^2 é chamado **comunalidade** (devido aos fatores) e ψ_i é chamado de **variância específica**.
- $\text{Cov}(Z_i, Z_k) = l_{i1}l_{1k} + \ldots + l_{im}l_{mk}$ para $i, k = 1, \ldots, p$ e $i \neq k$.
- $\text{Cov}(\underset{\sim}{Z}, \underset{\sim}{F}) = L$ e $\text{Cov}(Z_i, F_k) = \text{Cor}(Z_i, F_k) = l_{ik}$ pois $\text{Cov}(\underset{\sim}{Z}, \underset{\sim}{F}) = \text{Cov}(L\underset{\sim}{F} + \underset{\sim}{\epsilon}, \underset{\sim}{F}) = \text{Cov}(L\underset{\sim}{F}, \underset{\sim}{F}) + \text{Cov}(\underset{\sim}{\epsilon}, \underset{\sim}{F}) = L$.

Modelo fatorial ortogonal

Critérios para a escolha do número de fatores

- Análise da proporção da variância,
- Número de autovalores de R maiores do que 1,
- Gráfico scree-plot.

Modelo fatorial ortogonal

Pode-se estimar L e Ψ utilizando alguns métodos

1. Método dos fatores principais
2. Método dos fatores principais iterativo
3. Método da máxima verossimilhança (com a suposição de normalidade).

Métodos para a estimação das matrizes L e Ψ

1. Método das componentes principais ou fatores principais:

Considere a decomposição espectral de P:

$$\begin{aligned} P &= \sum_{i=1}^{p} \lambda_i \underset{\sim}{e}_i \underset{\sim}{e}_i^\top = \\ &= \begin{pmatrix} \sqrt{\lambda_1}\underset{\sim}{e}_1 & \dots & \sqrt{\lambda_p}\underset{\sim}{e}_p \end{pmatrix} \begin{pmatrix} \sqrt{\lambda_1}\underset{\sim}{e}_1^\top \\ \dots \\ \sqrt{\lambda_p}\underset{\sim}{e}_p^\top \end{pmatrix} = LL^\top \end{aligned}$$

Se o número de fatores $m = p$, então $P = LL^\top$. Entretanto, nosso interesse está quase sempre em reduzir a dimensionalidade do problema para $m < p$ fatores. Nesse caso, consideramos o algoritmo:

Métodos para a estimação das matrizes L e Ψ

1. Método das componentes principais ou fatores principais:

Considere a decomposição espectral de P:

$$\begin{aligned} P &= \sum_{i=1}^{p} \lambda_i \underset{\sim}{e}_i \underset{\sim}{e}_i^\top = \\ &= \begin{pmatrix} \sqrt{\lambda_1}\underset{\sim}{e}_1 & \dots & \sqrt{\lambda_p}\underset{\sim}{e}_p \end{pmatrix} \begin{pmatrix} \sqrt{\lambda_1}\underset{\sim}{e}_1^\top \\ \dots \\ \sqrt{\lambda_p}\underset{\sim}{e}_p^\top \end{pmatrix} = LL^\top \end{aligned}$$

Se o número de fatores $m = p$, então $P = LL^\top$. Entretanto, nosso interesse está quase sempre em reduzir a dimensionalidade do problema para $m < p$ fatores. Nesse caso, consideramos o algoritmo:

Método dos fatores principais

1. Extrair os autovalores e autovetores normalizados correspondentes de R, $(\widehat{\lambda}_i, \widehat{\underset{\sim}{e}}_i)$, para $i = 1, \dots, p$,
2. Selecionar os m autovalores e autovetores correspondentes $(\widehat{\lambda}_i, \widehat{\underset{\sim}{e}}_i)$, para $i = 1, \dots, m$,
3. Estimar L e Ψ fazendo:

$$\begin{aligned} \widehat{L} &= \left(\sqrt{\widehat{\lambda}_1}\widehat{\underset{\sim}{e}}_1, \dots, \sqrt{\widehat{\lambda}_m}\widehat{\underset{\sim}{e}}_m\right) \\ \widehat{\Psi} &= diag(R - \widehat{L}\widehat{L}^\top) \end{aligned}.$$

A matriz residual nesse caso é dada por

$$MRes = R - \widehat{L}\widehat{L}^\top - \widehat{\Psi}$$

Método dos fatores principais

1. Extrair os autovalores e autovetores normalizados correspondentes de R, $(\widehat{\lambda}_i, \widehat{\underset{\sim}{e}}_i)$, para $i = 1, \ldots, p$,
2. Selecionar os m autovalores e autovetores correspondentes $(\widehat{\lambda}_i, \widehat{\underset{\sim}{e}}_i)$, para $i = 1, \ldots, m$,
3. Estimar L e Ψ fazendo:

$$\begin{array}{rcl} \widehat{L} & = & \left(\sqrt{\widehat{\lambda}_1}\widehat{\underset{\sim}{e}}_1, \ldots, \sqrt{\widehat{\lambda}_m}\widehat{\underset{\sim}{e}}_m\right) \\ \widehat{\Psi} & = & diag(R - \widehat{L}\widehat{L}^\top) \end{array}.$$

A matriz residual nesse caso é dada por

$$MRes = R - \widehat{L}\widehat{L}^\top - \widehat{\Psi}$$

Métodos para a estimação das matrizes L e Ψ

2. Método das componentes principais iterativo ou fatores principais iterativo:

- Refinamento das estimativas de L e Ψ obtidas pelo Método dos fatores principais.
- m é definido por um critério anterior

Método dos fatores principais iterativo

Considere que a matriz de correlações de X, P, seja modelada como

$$P = LL^\top + \Psi$$

Então

$$LL^\top = P - \Psi = \begin{pmatrix} h_1^2 & \rho_{12} & \dots & \rho_{1p} \\ \rho_{12} & h_2^2 & \dots & \rho_{2p} \\ \vdots & \ddots & \ddots & \vdots \\ \rho_{p1} & \dots & \rho_{p-1,p} & h_p^2 \end{pmatrix}$$

em que $h_i^2 = 1 - \psi_i$, $i = 1, \dots, p$, são as comunalidades.

Método dos fatores principais iterativo

Suponha que se estime a matriz $LL^\top$ por $R^\star$ dada por

$$R^\star = \begin{pmatrix} h_1^{\star 2} & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{12} & h_2^{\star 2} & \cdots & \rho_{2p} \\ \vdots & \ddots & \ddots & \vdots \\ \rho_{p1} & \cdots & \rho_{p-1,p} & h_p^{\star 2} \end{pmatrix} \cong L^\star L^{\star\top}.$$

As quantidades $h_1^{\star 2}, \ldots, h_p^{\star 2}$ são as estimativas iniciais das comunalidades ${h_1}^2, \ldots, {h_p}^2$. Usando o método das componentes principais, temos

$$L^\star = (\sqrt{\widehat{\lambda}_1^\star}\widehat{\underset{\sim}{e}}_1^\star, \ldots, \sqrt{\widehat{\lambda}_m^\star}\widehat{\underset{\sim}{e}}_m^\star),$$

em que $\widehat{\lambda}_1^\star, \ldots, \widehat{\lambda}_m^\star$ são os autovalores da matriz $R^\star$ com autovetores correspondentes $\widehat{\underset{\sim}{e}}_1^\star, \ldots, \widehat{\underset{\sim}{e}}_m^\star$.

Método dos fatores principais iterativo

Suponha que se estime a matriz $LL^\top$ por $R^\star$ dada por

$$R^\star = \begin{pmatrix} h_1^{\star 2} & \rho_{12} & \dots & \rho_{1p} \\ \rho_{12} & h_2^{\star 2} & \dots & \rho_{2p} \\ \vdots & \ddots & \ddots & \vdots \\ \rho_{p1} & \dots & \rho_{p-1,p} & h_p^{\star 2} \end{pmatrix} \cong L^\star {L^\star}^\top.$$

As quantidades $h_1^{\star 2}, \dots, h_p^{\star 2}$ são as estimativas iniciais das comunalidades ${h_1}^2, \dots, {h_p}^2$. Usando o método das componentes principais, temos

$$L^\star = (\sqrt{\widehat{\lambda}_1^\star}\, \underset{\sim}{\widehat{e}}_1^\star, \dots, \sqrt{\widehat{\lambda}_m^\star}\, \underset{\sim}{\widehat{e}}_m^\star),$$

em que $\widehat{\lambda}_1^\star, \dots, \widehat{\lambda}_m^\star$ são os autovalores da matriz $R^\star$ com autovetores correspondentes $\underset{\sim}{\widehat{e}}_1^\star, \dots, \underset{\sim}{\widehat{e}}_m^\star$.

Método dos fatores principais iterativo

Suponha que se estime a matriz $LL^\top$ por $R^\star$ dada por

$$R^\star = \begin{pmatrix} h_1^{\star 2} & \rho_{12} & \dots & \rho_{1p} \\ \rho_{12} & h_2^{\star 2} & \dots & \rho_{2p} \\ \vdots & \ddots & \ddots & \vdots \\ \rho_{p1} & \dots & \rho_{p-1,p} & h_p^{\star 2} \end{pmatrix} \cong L^\star L^{\star\top}.$$

As quantidades $h_1^{\star 2}, \dots, h_p^{\star 2}$ são as estimativas iniciais das comunalidades ${h_1}^2, \dots, {h_p}^2$. Usando o método das componentes principais, temos

$$L^\star = (\sqrt{\widehat{\lambda}_1^\star}\,\widehat{\underset{\sim}{e}}{}_1^\star, \dots, \sqrt{\widehat{\lambda}_m^\star}\,\widehat{\underset{\sim}{e}}{}_m^\star),$$

em que $\widehat{\lambda}_1^\star, \dots, \widehat{\lambda}_m^\star$ são os autovalores da matriz $R^\star$ com autovetores correspondentes $\widehat{\underset{\sim}{e}}{}_1^\star, \dots, \widehat{\underset{\sim}{e}}{}_m^\star$.

Método dos fatores principais iterativo

A cada passo, as variâncias específicas são estimadas por

$$\psi_i^\star = 1 - {h_i^\star}^2.$$

A partir da matriz $L^\star$, temos as novas estimativas das comunalidades ${h_1}^2, \ldots, {h_p}^2$, que são então colocadas na diagonal principal da matriz $R^\star = L^\star {L^\star}^\top$ e o processo é repetido até que as diferenças nas estimativas das comunalidades de duas iterações sucessivas sejam suficientemente pequenas.

Como estimativas iniciais das comunalidades, podemos também considerar para h_i^2 o coeficiente de determinação R^2 do modelo de regressão linear em que Z_i é a variável resposta e as outras $p-1$ variáveis são variáveis explicativas.

Método dos fatores principais iterativo

A cada passo, as variâncias específicas são estimadas por

$$\psi_i^\star = 1 - h_i^{\star 2}.$$

A partir da matriz $L^\star$, temos as novas estimativas das comunalidades ${h_1}^2, \ldots, {h_p}^2$, que são então colocadas na diagonal principal da matriz $R^\star = L^\star {L^\star}^\top$ e o processo é repetido até que as diferenças nas estimativas das comunalidades de duas iterações sucessivas sejam suficientemente pequenas.

Como estimativas iniciais das comunalidades, podemos também considerar para h_i^2 o coeficiente de determinação R^2 do modelo de regressão linear em que Z_i é a variável resposta e as outras $p-1$ variáveis são variáveis explicativas.

Método dos fatores principais iterativo

A cada passo, as variâncias específicas são estimadas por

$$\psi_i^\star = 1 - h_i^{\star 2}.$$

A partir da matriz $L^\star$, temos as novas estimativas das comunalidades ${h_1}^2, \ldots, {h_p}^2$, que são então colocadas na diagonal principal da matriz $R^\star = L^\star {L^\star}^\top$ e o processo é repetido até que as diferenças nas estimativas das comunalidades de duas iterações sucessivas sejam suficientemente pequenas.

Como estimativas iniciais das comunalidades, podemos também considerar para h_i^2 o coeficiente de determinação R^2 do modelo de regressão linear em que Z_i é a variável resposta e as outras $p-1$ variáveis são variáveis explicativas.

Métodos para a estimação das matrizes L e Ψ

3. Método da máxima verossimilhança:

Suponha que $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$ e portanto $\underset{\sim}{Z} \sim N_p(\underset{\sim}{0}, P)$. Além disso, $\underset{\sim}{F} \sim N_p(\underset{\sim}{0}, I)$ e $\underset{\sim}{\epsilon} \sim N_p(\underset{\sim}{0}, \Psi)$.

A função de verossimilhança considerando uma amostra aleatória de tamanho n, $\underset{\sim}{Z}_1, \ldots, \underset{\sim}{Z_n}$, é expressa por

$$\mathscr{L}(P) = \frac{1}{(2\pi)^{np/2}|P|^{n/2}} \exp\{-\frac{1}{2}\sum_{j=1}^{n} \underset{\sim}{z_j}^{\top} P^{-1} \underset{\sim}{z}_j\}$$

Método da máxima verossimilhança

ou seja,

$$\mathscr{L}(L,\Psi)=\frac{1}{(2\pi)^{np/2}|LL^\top+\Psi|^{n/2}}\exp\{-\frac{1}{2}\sum_{j=1}^{n}\underset{\sim}{z_j}^\top(LL^\top+\Psi)^{-1}\underset{\sim}{z_j}\}$$

e devemos encontrar o máximo de $\mathscr{L}(L,\Psi)$ em L e Ψ. Essa maximização é feita em termos numéricos para um valor de m fixo.

Problema: o algoritmo pode não convergir.

Método da máxima verossimilhança

ou seja,

$$\mathscr{L}(L,\Psi) = \frac{1}{(2\pi)^{np/2}|LL^\top + \Psi|^{n/2}} \exp\{-\frac{1}{2}\sum_{j=1}^{n} \underset{\sim}{z}_j^\top (LL^\top + \Psi)^{-1} \underset{\sim}{z}_j\}$$

e devemos encontrar o máximo de $\mathscr{L}(L,\Psi)$ em L e Ψ. Essa maximização é feita em termos numéricos para um valor de m fixo.

Problema: o algoritmo pode não convergir.

Métodos para a estimação das matrizes L e Ψ

Quando usar um método ou outro?

Se a suposição de normalidade é válida, o método da máxima verossimilhança pode produzir estimativas mais precisas.

Se não podemos supor a normalidade multivariada, convém utilizar o método de fatores principais ou fatores principais iterativo.

Observação: Diferentes valores de m podem levar a diferentes estimativas das cargas fatoriais.

Métodos para a estimação das matrizes L e Ψ

Quando usar um método ou outro?

Se a suposição de normalidade é válida, o método da máxima verossimilhança pode produzir estimativas mais precisas.

Se não podemos supor a normalidade multivariada, convém utilizar o método de fatores principais ou fatores principais iterativo.

Observação: Diferentes valores de m podem levar a diferentes estimativas das cargas fatoriais.

Métodos para a estimação das matrizes L e Ψ

Quando usar um método ou outro?

Se a suposição de normalidade é válida, o método da máxima verossimilhança pode produzir estimativas mais precisas.

Se não podemos supor a normalidade multivariada, convém utilizar o método de fatores principais ou fatores principais iterativo.

Observação: Diferentes valores de m podem levar a diferentes estimativas das cargas fatoriais.

Métodos para a estimação das matrizes L e Ψ

Quando usar um método ou outro?

Se a suposição de normalidade é válida, o método da máxima verossimilhança pode produzir estimativas mais precisas.

Se não podemos supor a normalidade multivariada, convém utilizar o método de fatores principais ou fatores principais iterativo.

Observação: Diferentes valores de m podem levar a diferentes estimativas das cargas fatoriais.

Rotação ortogonal de fatores (Mingoti, 2007)

Objetivo: Aplicar transformações ortogonais aos fatores originais de modo que os novos fatores obtidos tenham interpretações mais fáceis e diretas.

Ideia: Considerar a matriz T ortogonal tal que $TT^\top = T^\top T = I$ e rotacionar a matriz de cargas fatoriais fazendo

$$\widehat{L}^\star = \widehat{L}T.$$

Note que

$$\widehat{L}^\star \widehat{L}^{\star\top} = (\widehat{L}T)(\widehat{L}T)^\top = \widehat{L}\widehat{L}^\top.$$

Rotação ortogonal de fatores (Mingoti, 2007)

Objetivo: Aplicar transformações ortogonais aos fatores originais de modo que os novos fatores obtidos tenham interpretações mais fáceis e diretas.

Ideia: Considerar a matriz T ortogonal tal que $TT^\top = T^\top T = I$ e rotacionar a matriz de cargas fatoriais fazendo

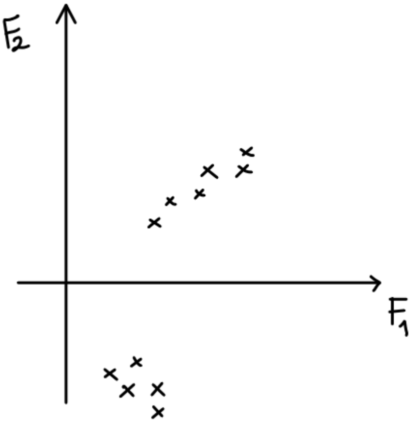
$$\widehat{L}^\star = \widehat{L}T.$$

Note que

$$\widehat{L}^\star \widehat{L}^{\star\top} = (\widehat{L}T)(\widehat{L}T)^\top = \widehat{L}\widehat{L}^\top.$$

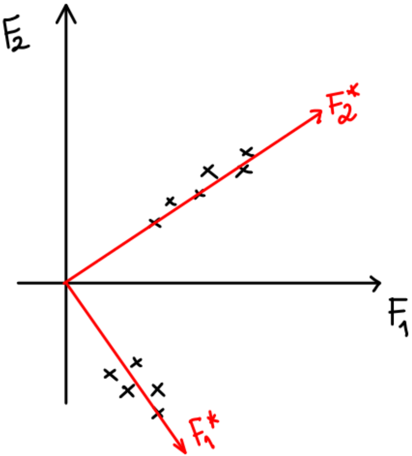
Rotação ortogonal de fatores

Interpretação geométrica $(p = 2)$



Rotação ortogonal de fatores

Interpretação geométrica $(p = 2)$



Critérios para a rotação ortogonal de fatores

Critério varimax: (Kaiser 1958)

Objetivo: Encontrar fatores com variabilidade máxima nas cargas fatoriais.

Seja $\hat{l}^{\star}_{ij}$ o coeficiente da i-ésima variável no j-ésimo fator após a rotação.

Seja

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

em que $\tilde{l}_{ij} = \dfrac{\hat{l}^{\star}_{ij}}{\hat{h}_i}$.

O critério **varimax** busca $\tilde{l}_{ij}$ que levam a V máximo para $i = 1, \ldots, p$.

Critérios para a rotação ortogonal de fatores

Critério quartimax: (Jobson, 1996)

Objetivo: Encontrar fatores que levem ao máximo da variabilidade dos quadrados das cargas fatoriais sobre todos os fatores.

Seja $\hat{l}_{ij}^{\star}$ o coeficiente da i-ésima variável no j-ésimo fator após a rotação.

Seja

$$V_Q = \frac{1}{pm} \sum_{j=1}^{m} \sum_{i=1}^{p} \hat{l}_{ij}^{\star 4} - \frac{1}{pm} \sum_{j=1}^{m} \sum_{i=1}^{p} (\hat{l}_{ij}^{\star 2})^2$$

O critério **quartimax** busca $\tilde{l}_{ij}$ que levam a V_Q ao seu valor máximo.

Critérios para a rotação ortogonal de fatores

Critério orthomax: (Jobson, 1996)

Seja $\hat{l}^{\star}_{ij}$ o coeficiente da i-ésima variável no j-ésimo fator após a rotação.

Seja

$$V_M = \frac{1}{pm} \sum_{j=1}^{m} \left[\sum_{i=1}^{p} \hat{l}^{\star 4}_{ij} - \frac{\gamma}{p} \sum_{i=1}^{m} \sum_{i=1}^{p} (\hat{l}^{\star 2}_{ij})^2 \right]$$

O critério **orthomax** busca $\hat{l}^{\star}_{ij}$ que levam a V_M ao seu valor máximo.

Se $\gamma = 0$, temos o critério quartimax.

Se $\gamma = 1$, o critério se assemelha ao varimax.

Se $\gamma = 0,5$, o critério se chama biquartimax.

Estimação dos escores $\widehat{F}$ para cada elemento amostral

Os escores das componentes podem ser usados para a ordenação das observações, por exemplo. Duas possíveis estratégias:

Mínimos quadrados ponderados

$$\widehat{\underset{\sim}{F}}_j = (\widehat{L}^\top \widehat{\Psi}^{-1} \widehat{L})^{-1} (\widehat{L}^\top \widehat{\Psi}^{-1}) \underset{\sim}{z}_j, j = 1, \ldots, n$$

Método de regressão

$$\widehat{\underset{\sim}{F}}_j = \widehat{L}^\top \widehat{P}^{-1} \underset{\sim}{z}_j, j = 1, \ldots, n$$

Estimação dos escores $\widehat{F}$ para cada elemento amostral

Os escores das componentes podem ser usados para a ordenação das observações, por exemplo. Duas possíveis estratégias:

Mínimos quadrados ponderados

$$\widehat{\underset{\sim}{F}}_j = (\widehat{L}^\top \widehat{\Psi}^{-1} \widehat{L})^{-1} (\widehat{L}^\top \widehat{\Psi}^{-1}) \underset{\sim}{z}_j, j = 1, \ldots, n$$

Método de regressão

$$\widehat{\underset{\sim}{F}}_j = \widehat{L}^\top \widehat{P}^{-1} \underset{\sim}{z}_j, j = 1, \ldots, n$$