Análise Fatorial

Prof. Caio Azevedo

Motivação

- Charles Spearman, Thomson, Thurstone e Burt, buscaram obter uma melhor compreensão para "inteligência". Este conceito está relacionado à variáveis cognitivas.
- Eles conjecturaram a possibilidade de que os desempenhos de indivíduos (habilidades) em testes cognitivos, em diferentes áreas, pudessem ser representados por um número menor de variáveis não observáveis (latentes).
- Colocando de uma outra forma, tais variáveis latentes podem ser mensuradas através do desempenho dos indivíduos.

Exemplo 5: Johnson & Wichern

- Notas de indivíduos em diferentes testes. Matriz de correlações baseada nos resultados de 220 indivíduos. Análise: dados simulados a partir da matriz de correlações.
- Variáveis:
  - Gaelic: idioma gaélico (Ga).
  - English: língua inglesa (I).
  - History: História (H).
  - Arithmetic: Aritmética (Ar).
  - Algebra: Álgebra (Al).
  - Geometry: Geometria (Ge).

Matriz de correlações

	Ga	I	H	Ar	Al	Ge
Ga	1,000	0,439	0,410	0,288	0,329	0,248
I	0,439	1,000	0,351	0,354	0,320	0,329
H	0,410	0,351	1,000	0,164	0,190	0,181
Ar	0,288	0,354	0,164	1,000	0,595	0,470
Al	0,329	0,320	0,190	0,595	1,000	0,464
Ge	0,248	0,329	0,181	0,470	0,464	1,000

Dados simulados ($N_6(\boldsymbol{\mu}, \boldsymbol{\Sigma})$)

- Médias e variâncias artificiais: $\boldsymbol{\mu} = (8,2; 8,0; 7,6; 9,1; 9,4; 8,9)'$ e $\psi = (1,0; 1,5; 0,8; 1,3; 1,7; 2)'$.
- Matriz de covariâncias ($\boldsymbol{\Sigma}$). Seja $\mathbf{R}$ a matriz de correlações anterior e $\boldsymbol{\Psi} = \text{diag}(\sqrt{1,0}; \sqrt{1,5}; \sqrt{0,8}; \sqrt{1,3}; \sqrt{1,7}; \sqrt{2,0})$. Então $\boldsymbol{\Sigma} = \boldsymbol{\Psi}\mathbf{R}\boldsymbol{\Psi}$.
- Na matriz de dados simulados, digamos $\mathbf{X}$, foi realizada a seguinte transformação (a fim de assegurar que a matriz de covariâncias, portanto a de correlações, amostrais, sejam muito próximas das desejadas): $\mathbf{X}^* = (\mathbf{X} \times (Chol(\mathbf{S}^2))^{-1}) \times Chol(\boldsymbol{\Psi})$, em que $\mathbf{S}^2$ é a matriz de covariâncias amostrais de $\mathbf{X}$.

Modelo de análise fatorial (MAF)

- Temos:

$$
\begin{aligned}
X_1 &= \mu_1 + l_{11}F_1 + l_{12}F_2 + ... + l_{1m}F_m + \xi_1 \\
X_2 &= \mu_2 + l_{21}F_1 + l_{22}F_2 + ... + l_{2m}F_m + \xi_2 \\
\vdots &= \qquad\qquad\qquad \vdots \\
X_p &= \mu_p + l_{p1}F_1 + l_{p2}F_2 + ... + l_{pm}F_m + \xi_p
\end{aligned}
$$

- $F_i, i = 1, 2, .., m$, são os fatores (ou fatores comuns), $\xi_i, i = 1, ..., p$ são erros aleatórios (ou fatores específicos) e $l_{ij}, i = 1, .., p$; $j = 1, .., m$ são as chamadas cargas fatoriais (m $<$ p).
- Várias estruturas podem ser consideradas para F_i e ξ_i. Veremos o chamado modelo de análise fatorial ortogonal (MAFO).

MAFO (forma matricial)

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

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}; \mathbf{F} = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{bmatrix}; \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}; \mathbf{L} = \begin{bmatrix} l_{11} & l_{12} & \dots & l_{1m} \\ l_{21} & l_{22} & \dots & l_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ l_{p1} & l_{p2} & \dots & l_{pm} \end{bmatrix};$$

$$\boldsymbol{\xi} = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_p \end{bmatrix}.$$

MAFO (forma matricial)

Adicionalmente, $\mathcal{E}(\mathbf{F}) = \mathbf{0}_m$, $\mathcal{E}(\boldsymbol{\xi}) = \mathbf{0}_p$, $Cov(\mathbf{F}) = \mathbf{I}_m$, $Cov(\boldsymbol{\xi}) = \boldsymbol{\Psi}$ e $Cov(\boldsymbol{\xi}, \mathbf{F}) = \mathbf{0}_{(p \times m)}$, em que $\boldsymbol{\Psi} = \begin{bmatrix} \psi_1 & 0 & \ldots & 0 \\ 0 & \psi_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \psi_p \end{bmatrix}$.

MAFO (forma matricial)

- Assim, o MAF (e, particularmente o MAFO) é um modelo de regressão linear multivariado no qual as variáveis explicativas $(\mathbf{F})$ não são observáveis.
- Com efeito, os fatores são considerados variáveis aleatórias.
- É chamado de ortogonal pois está-se admitindo que os fatores comuns são não correlacionados $(Cov(\mathbf{F}) = \mathbf{I}_m)$.

MAFO: características e propriedades

- $\mathbf{F}$: vetor de fatores (comuns).
- $\boldsymbol{\xi}$: vetor de fatores específicos (ou erros aleatórios).
- $\mathbf{L}$: matriz de cargas fatoriais.
- $\mathcal{E}(\mathbf{X}) = \boldsymbol{\mu} + \mathcal{E}(\mathbf{F}) + \mathcal{E}(\boldsymbol{\xi}) = \boldsymbol{\mu} + \mathbf{0} + \mathbf{0} = \boldsymbol{\mu}$.
- $Cov(\mathbf{X}) = Cov(\mathbf{LF}) + Cov(\boldsymbol{\xi}) + Cov(\mathbf{LF}, \boldsymbol{\xi}) =$

  $\mathbf{L}Cov(\mathbf{F})\mathbf{L}' + \boldsymbol{\Psi} + \mathbf{0} = \mathbf{LL}' + \boldsymbol{\Psi}$.
- $Cov(\mathbf{X}, \mathbf{F}) = Cov(\boldsymbol{\mu} + \mathbf{LF} + \boldsymbol{\xi}, \mathbf{F}) =$

  $Cov(\boldsymbol{\mu}, \mathbf{F}) + Cov(\mathbf{LF}, \mathbf{F}) + Cov(\boldsymbol{\xi}, \mathbf{F}) = \mathbf{0} + \mathbf{L}Cov(\mathbf{F}, \mathbf{F}) + \mathbf{0} = \mathbf{L}$.
- Objetivo: estimar $\mathbf{L}$ (quantidade não aleatória) e predizer $\mathbf{F}$ (quantidade aleatória) com base em uma matriz de dados.

MAFO: propriedades

- Note que

$$
\mathbf{L}\mathbf{L}' = \begin{bmatrix} l_{11} & l_{12} & \ldots & l_{1m} \\ l_{21} & l_{22} & \ldots & l_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ l_{p1} & l_{p2} & \ldots & l_{pm} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & \ldots & l_{p1} \\ l_{12} & l_{22} & \ldots & l_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ l_{1m} & l_{2m} & \ldots & l_{pm} \end{bmatrix} =
$$

$$
\begin{bmatrix} \sum_{j=1}^{m} l_{1j}^2 & \sum_{j=1}^{m} l_{1j} l_{2j} & \ldots & \sum_{j=1}^{m} l_{1j} l_{pj} \\ \sum_{j=1}^{m} l_{1j} l_{2j} & \sum_{j=1}^{m} l_{2j}^2 & \ldots & \sum_{j=1}^{m} l_{2j} l_{pj} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^{m} l_{1j} l_{pj} & \sum_{j=1}^{m} l_{2j} l_{pj} & \ldots & \sum_{j=1}^{m} l_{pj}^2 \end{bmatrix}
$$

MAFO: propriedades

- Variância:

$$\mathcal{V}(X_i) = \underbrace{\sum_{j=1}^{m} l_{ij}^2}_{\text{comunalidade}} + \underbrace{\psi_i}_{\text{especificidade}} . \qquad (1)$$

- $Cov(X_i, X_j) = \sum_{k=1}^{m} l_{ik} l_{jk}$.
- Seja $T_{(m \times m)}$ uma matriz ortogonal $(\mathbf{T}^{-1} = \mathbf{T}')$ e defina $\mathbf{L}^* = \mathbf{LT}$, assim:

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

- Assim, note que qualquer transformação ortogonal na matriz de cargas fatoriais leva à mesma representação da matriz de covariâncias.

MAFO: propriedades

- Isso pode levar à um problema de falta de identificabilidade se, por exemplo, estimarmos as quantidades de interesse utilizando métodos baseados na verossimilhança.
- Resultado: Se $\mathbf{Y}|\mathbf{X} = \mathbf{x} \sim N_p(\mathbf{A}\mathbf{x} + \mathbf{B}, \boldsymbol{\Sigma})$ e $\mathbf{X} \sim N_q(\boldsymbol{\mu}, \boldsymbol{\Psi})$, então $\mathbf{Y} \sim N_p(\mathbf{A}\boldsymbol{\mu} + \mathbf{B}, \mathbf{A}\boldsymbol{\Psi}\mathbf{A}' + \boldsymbol{\Sigma})$.
- Exemplo: Se $\mathbf{F} \sim N_m(\mathbf{0}, \mathbf{I})$ e $\boldsymbol{\xi} \sim N_p(\mathbf{0}, \boldsymbol{\Psi})$ então $\mathbf{X}|\mathbf{F} = \mathbf{f} \sim N_p(\boldsymbol{\mu} + \mathbf{L}\mathbf{f}, \boldsymbol{\Psi})$. Assim, $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi})$

MAFO: propriedades

- Seja $L(\mathbf{L},\mathbf{\Psi}) = f(\mathbf{x})$, então

$$L(\mathbf{L},\mathbf{\Psi}) \equiv f_{\mathbf{x}}(\mathbf{x}) \propto |\mathbf{L}\mathbf{L}' + \mathbf{\Psi}|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\left(\mathbf{L}\mathbf{L}' + \mathbf{\Psi}\right)^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

$$L(\mathbf{L},\mathbf{\Psi}) = L(\mathbf{L}^*,\mathbf{\Psi})$$

- Falta de identificabilidade: diferentes valores dos parâmetros levam ao mesmo valores da verossimilhança.
- Tal aspecto pode ter de ser levado em consideração, consoante o mecanismo de estimação adotado para as cargas fatoriais.

Análise fatorial $\times$ Análise de componentes principais

Análise fatorial	**Análise de componentes principais**
Modelo estatístico	Sem modelo estatístico
Impõe uma estrutura específica para matriz de covariâncas dos dados	Não impõe estrutura
Busca, por definição, diminuir a dimensionalidade dos dados	A diminuição da dimensionalidade é uma consequência

Estimação

- Seja uma matriz de dados $\mathbf{X}_{(n \times p)}$

Indivíduo	Variável 1	Variável 2	...	Variável p
1	X_{11}	X_{12}	...	X_{1p}
2	X_{21}	X_{22}	...	X_{2p}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
n	X_{n1}	X_{n2}	...	X_{np}

- Estima-se as cargas fatoriais e, posteriormente, prediz-se os fatores.

MAFO para uma matriz de dados

$$
\begin{aligned}
X_{i1} &= \mu_1 + l_{11}F_{i1} + l_{12}F_{i2} + ... + l_{1m}F_{im} + \xi_{i1} \\
X_{i2} &= \mu_2 + l_{21}F_{i1} + l_{22}F_{i2} + ... + l_{2m}F_{im} + \xi_{i2} \\
\vdots &= \vdots \qquad\qquad\qquad \vdots \\
X_{ip} &= \mu_p + l_{p1}F_{i1} + l_{p2}F_{i2} + ... + l_{pm}F_{im} + \xi_{ip}
\end{aligned}
$$

i=1,2,...,n.

MAFO (matriz de dados)

$$\mathbf{X}_{(n\times p)} = \mathbf{1}_n\boldsymbol{\mu}' + \mathbf{F}_{(n\times m)}\mathbf{L}'_{(m\times p)} + \boldsymbol{\xi}_{(n\times p)}$$

- $\mathbf{X}$: matriz de dados.
- $\boldsymbol{\mu}$: vetor de médias e $\mathbf{1}_n$ é um vetor de 1's de tamanho n.
- $\mathbf{F}$: matriz de fatores (comuns), desconhecida e aleatória.
- $\mathbf{L}$: matriz de cargas fatoriais, desconhecida e não aleatória.
- $\boldsymbol{\xi}$: matriz de resíduos.
- $\mathcal{E}(\mathbf{F}_i) = \mathbf{0}_m$, $\mathcal{E}(\boldsymbol{\xi}_i) = \mathbf{0}_p$, $Cov(\mathbf{F}_i) = \mathbf{I}_m$, $Cov(\boldsymbol{\xi}_i) = \boldsymbol{\Psi}$ e $Cov(\boldsymbol{\xi}_i, \mathbf{F}_j) = \mathbf{0}_{(p\times m)}, \forall i,j$, em que $\boldsymbol{\Psi} = \text{diag}(\psi_1, ..., \psi_p)$, i,j=1,2,...,n.

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1p} \\ X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \dots & X_{np} \end{bmatrix} ; \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} ; \mathbf{L} = \begin{bmatrix} l_{11} & l_{12} & \ldots & l_{1m} \\ l_{21} & l_{22} & \ldots & l_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ l_{p1} & l_{p2} & \ldots & l_{pm} \end{bmatrix}$$

$$\mathbf{F} = \begin{bmatrix} F_{11} & F_{12} & \dots & F_{1m} \\ F_{21} & F_{22} & \dots & F_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ F_{n1} & F_{n2} & \dots & F_{nm} \end{bmatrix} = \begin{bmatrix} \mathbf{F}_1 \\ \mathbf{F}_2 \\ \vdots \\ \mathbf{F}_n \end{bmatrix} ; \boldsymbol{\xi} =$$

$$\begin{bmatrix} \xi_{11} & \xi_{12} & \dots & \xi_{1p} \\ \xi_{21} & \xi_{22} & \dots & \xi_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \xi_{n1} & \xi_{n2} & \dots & \xi_{np} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \\ \vdots \\ \boldsymbol{\xi}_n \end{bmatrix} ;$$

Estimação de **L**: método das componentes principais

- Pela decomposição espectral, temos que: $\mathbf{\Sigma} = \mathbf{E}\Lambda\mathbf{E}'$ (os autovetores ortonormalizados correspondem às colunas da matriz $\mathbf{E}$), em que

$$\mathbf{E} = \begin{bmatrix} e_{11} & e_{21} & \dots & e_{p1} \\ e_{12} & e_{22} & \dots & e_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ e_{1p} & e_{p2} & \dots & e_{pp} \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \dots & \mathbf{e}_p \end{bmatrix}$$

Estimação de $\mathbf{L}$: método das componentes principais

- Suponha um modelo com p fatores e sem contribuição dos erros aleatórios (fatores específicos), ou seja, $\psi_i = 0, i = 1, 2, ..., p$, assim, teríamos

$$
\begin{aligned}
\boldsymbol{\Sigma} &= \mathbf{E}\Lambda\mathbf{E}' = (\mathbf{E}\boldsymbol{\Lambda}^{1/2})(\boldsymbol{\Lambda}^{1/2}\mathbf{E}') = \mathbf{L}\mathbf{L}' + \underbrace{\boldsymbol{\Psi}}_{\mathbf{0}} \\
&\rightarrow \begin{bmatrix} \sqrt{\lambda_1}\mathbf{e}_1 & \sqrt{\lambda_2}\mathbf{e}_2 & \ldots & \sqrt{\lambda_p}\mathbf{e}_p \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1}\mathbf{e}_1' \\ \sqrt{\lambda_2}\mathbf{e}_2' \\ \ldots \\ \sqrt{\lambda_p}\mathbf{e}_p' \end{bmatrix} = \mathbf{L}\mathbf{L}'
\end{aligned}
$$

Estimação de $\mathbf{L}$: método das componentes principais

Logo

$$
\begin{aligned}
\mathbf{L} &= \begin{bmatrix} \sqrt{\lambda_1}\mathbf{e}_1 & \sqrt{\lambda_2}\mathbf{e}_2 & \dots & \sqrt{\lambda_p}\mathbf{e}_p \end{bmatrix} \\
&\rightarrow \begin{bmatrix} l_{11} & l_{12} & \dots & l_{1p} \\ l_{21} & l_{22} & \dots & l_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ l_{p1} & l_{p2} & \dots & l_{pp} \end{bmatrix} = \begin{bmatrix} \sqrt{\lambda_1}e_{11} & \sqrt{\lambda_2}e_{21} & \dots & \sqrt{\lambda_p}e_{p1} \\ \sqrt{\lambda_1}e_{12} & \sqrt{\lambda_2}e_{22} & \dots & \sqrt{\lambda_p}e_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{\lambda_1}e_{1p} & \sqrt{\lambda_2}e_{p2} & \dots & \sqrt{\lambda_p}e_{pp} \end{bmatrix}
\end{aligned}
$$

Assim $l_{ij} = \sqrt{\lambda_i}e_{ij}$. Na prática, utilizamos a matriz de variâncias-covariâncias amostrais, assim $\widetilde{l}_{ij} = \sqrt{\widetilde{\lambda}_j}\widetilde{e}_{ji}$.

Estimação de $\mathbf{L}$: método das componentes principais

- Suponha que agora que m $<$ p (m fatores), assim $\boldsymbol{\Sigma} = \mathbf{E}\Lambda\mathbf{E}' + \boldsymbol{\Psi}$, logo (veja equação (1)), uma estimativa para $\psi_i, i = 1, 2, .., p$ é dada por

$$\widetilde{\psi}_i = \widetilde{\sigma}_i^2 - \sum_{j=1}^{m} \widetilde{l}_{ij}^2$$

em que $\widetilde{\sigma}_i^2$ é a variância amostral da variável i.
- Para estimar $\boldsymbol{\mu}$ utilizamos o vetor de médias amostrais ($\overline{\mathbf{x}}$), como visto anteriormente.
- Naturalmente desejamos que $\widetilde{\boldsymbol{\Sigma}} \approx \widetilde{\mathbf{L}}\widetilde{\mathbf{L}}' + \widetilde{\boldsymbol{\Psi}}$.
- Neste procedimento não há problemas de falta de identificabilidade.

Observações

- De (1), temos que a contribuição do fator j para explicar a variância da variável i é dada por l_{ij}^2.
- Assim, tal contribuição, para explicar a soma das variâncias de todas as variáveis (variância total) é dada por $\sum_{i=1}^{p} \widetilde{l}_{ij}^2$.
- Pode-se provar que $\sum_{i=1}^{p} \widetilde{l}_{ij}^2 = \widetilde{\lambda}_j$.

Observações

- Assim, um outro critério que pode auxiliar na escolha do número de fatores a serem considerados é a proporção da variância total explicada por cada fator (e acumulada), ou seja $\frac{\widetilde{\lambda}_j}{\sum_{i=1}^{p}\widetilde{\sigma}_i^2}$ (PVE pelo fator j).
- Os resultados também se aplicam se trabalharmos com variáveis com variância unitária, ou seja, se usarmos $\boldsymbol{\rho}$ ao invés de $\boldsymbol{\Sigma}$.
- Nesse caso $(\boldsymbol{\rho})$:

$$\widetilde{\psi}_i = 1 - \sum_{j=1}^{m} \widetilde{l}_{ij}^2; PVE_j = \frac{\widetilde{\lambda}_j}{p}$$

Estimação de $\mathbf{L}$ e $\boldsymbol{\Psi}$: método de máxima verossimilhança

- Maximizar

$$L(\mathbf{L}, \boldsymbol{\Psi}) \propto |\mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}|^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})'\left(\mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}\right)^{-1}(\mathbf{x}_i - \boldsymbol{\mu})\right\}$$

em relação à $\mathbf{L}$ e $\boldsymbol{\Psi}$.

- Problema: falta de identificabilidade.
- A maximização analítica é complicada. Algum método númerico de otimização tem de ser utilizado (Newton-Raphson, Escore de Fisher, BFGS, Nelder-Mead).

Identificabilidade

- Fixar alguns valores de $\mathbf{L}$ ou de $\mathbf{LL}'$.
- Restrição: $\mathbf{L}'\boldsymbol{\Psi}^{-1}\mathbf{L} = \boldsymbol{\Delta}$, em que $\boldsymbol{\Delta}$ é uma matriz diagonal. Note que a verossimilhança depende de $\boldsymbol{\Psi}$ e $\mathbf{L}$ somente através de $\left(\boldsymbol{\Psi} + \mathbf{LL}'\right)^{-1}$. Além disso:

$$
\begin{aligned}
\left(\boldsymbol{\Psi} + \mathbf{LL}'\right)^{-1} &= \boldsymbol{\Psi}^{-1} - \boldsymbol{\Psi}^{-1}\mathbf{L}\left(\mathbf{I} + \mathbf{L}'\boldsymbol{\Psi}^{-1}\mathbf{L}\right)^{-1}\mathbf{L}'\boldsymbol{\Psi}^{-1} \\
&= \boldsymbol{\Psi}^{-1}\left[\mathbf{I} - \mathbf{L}\left(\mathbf{I} + \underbrace{\mathbf{L}'\boldsymbol{\Psi}^{-1}\mathbf{L}}_{\boldsymbol{\Delta}}\right)^{-1}\mathbf{L}'\boldsymbol{\Psi}^{-1}\right]
\end{aligned}
$$

Cont.

- Note que (sendo $\mathbf{T}$ uma matriz ortogonal):

$$
\begin{aligned}
\mathbf{L}(\mathbf{I}+\boldsymbol{\Delta})^{-1}\mathbf{L}' &= \mathbf{L}\mathbf{T}\mathbf{T}'(\mathbf{I}+\boldsymbol{\Delta})^{-1}(\mathbf{L}\mathbf{T}\mathbf{T}')' \\
&= \mathbf{L}\mathbf{T}\mathbf{T}^{-1}(\mathbf{I}+\boldsymbol{\Delta})^{-1}(\mathbf{T}^{-1})^{-1}\mathbf{T}'\mathbf{L}' \\
&= \mathbf{L}\mathbf{T}\left(\mathbf{T}\mathbf{T}^{-1}+\mathbf{T}\boldsymbol{\Delta}\mathbf{T}^{-1}\right)^{-1}\mathbf{T}'\mathbf{L}' = \mathbf{L}\mathbf{T}\left(\mathbf{I}+\mathbf{T}\boldsymbol{\Delta}\mathbf{T}'\right)^{-1}\mathbf{T}'\mathbf{L}' \\
&= \mathbf{L}^{*}\left(\mathbf{I}+\mathbf{T}\boldsymbol{\Delta}\mathbf{T}'\right)^{-1}(\mathbf{L}^{*})'
\end{aligned}
$$

em que $\mathbf{L}^{*} = \mathbf{L}\mathbf{T}$ e $\mathbf{T}\boldsymbol{\Delta}\mathbf{T}' \neq \boldsymbol{\Delta}$.

Determinação do número de fatores

- Desejamos que $\widetilde{\boldsymbol{\Sigma}} \approx \widetilde{\mathbf{L}}\widetilde{\mathbf{L}}^{'} + \widetilde{\boldsymbol{\Psi}}$.
- Um outro critério que pode auxiliar na escolha do número de fatores a serem considerados é a proporção da variância total explicada por cada fator (e acumulada), ou seja $\frac{\sum_{i=1}^{p} \widetilde{l}_{ij}^{2}}{\sum_{i=1}^{p} \widetilde{\sigma}_{i}^{2}}$ (PVE pelo fator i).
- Se trabalharmos com variáveis com variância unitária, ou seja, se usarmos $\boldsymbol{\rho}$ ao invés de $\boldsymbol{\Sigma}$, teremos $PVE_i = \frac{\sum_{i=1}^{p} \widetilde{l}_{ij}^{2}}{p}$
- Testes de hipótese.

Testes de hipótese para determinação do número de fatores via MV

- O resultado é válido se considerarmos $\boldsymbol{\Sigma}$ ou $\boldsymbol{\rho}$
- Desejamos testar se $H_0 : \boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}$ vs $H_1 : \boldsymbol{\Sigma} \neq \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}$. Seja $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (irrestrito), $\boldsymbol{\theta}_0 = (\boldsymbol{\mu}, \mathbf{L}, \boldsymbol{\Psi})$ (restrito).
- Teste da razão de verossimilhanças: $\boldsymbol{\Theta}_0$ - espaço paramétrico sob H_0 $\boldsymbol{\Theta}$ - espaço paramétrico irrestrito.
- Estatística: $\Lambda = \frac{sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} L(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} L(\boldsymbol{\mu}, \boldsymbol{\Sigma})} = \frac{sup_{\boldsymbol{\theta}_0} L(\boldsymbol{\mu}, \mathbf{L}, \boldsymbol{\Psi})}{sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} L(\boldsymbol{\mu}, \boldsymbol{\Sigma})}$.
- Sob H_0 ou irrestritamente o emv de $\boldsymbol{\mu}$ é dado por $\overline{\mathbf{X}}$.

Testes de hipótese para determinação do número de fatores via MV

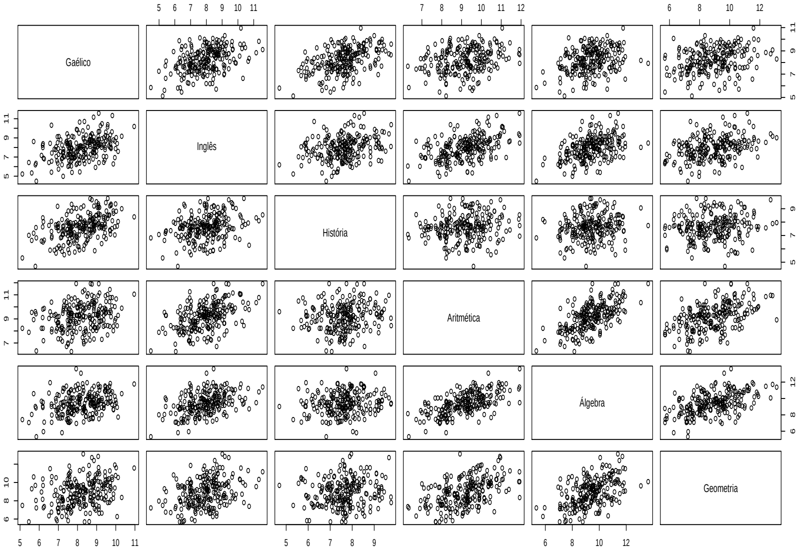
- Irrestritamente o emv de $\boldsymbol{\Sigma}$ é dado por $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{x}_i - \overline{\mathbf{x}}\right)\left(\mathbf{x}_i - \overline{\mathbf{x}}\right)'$.
- Os estimadores de MV de $\mathbf{L}$ e $\boldsymbol{\Psi}$ são obtidos numericamente.
- É bastante complicado obter a distribuição exata (mesmo usando aproximações analíticas) de Λ ou mesmo $\lambda = -2\ln\Lambda$.
- Contudo, sob H_0, $\lambda \approx \chi^2_{(r)}$, para n suficientemente grande.
- Vamos determinar r.

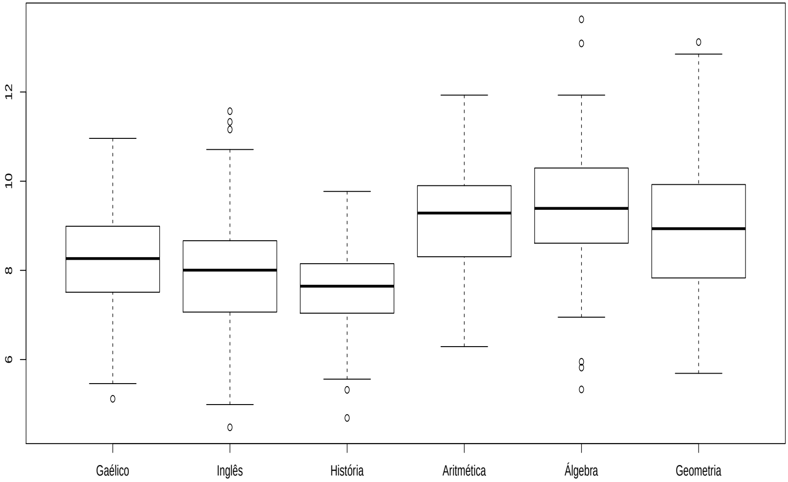
Testes de hipótese para determinação do número de fatores via MV

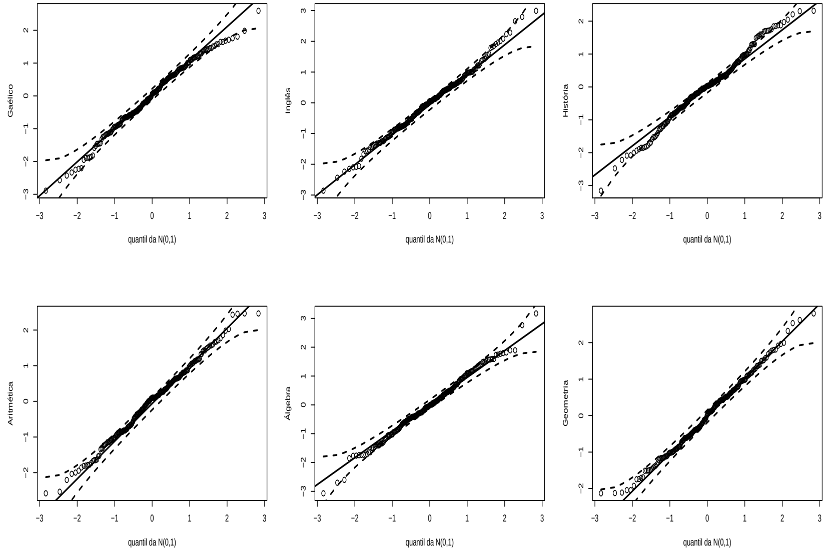
- $\mathbf{\Sigma}$ tem $p(p+1)/2$ parâmetros, $\mathbf{L}$ tem pm parâmetros e $\mathbf{\Psi}$ m. Contudo, devido à restrição $\mathbf{L}'\mathbf{\Psi}^{-1}\mathbf{L} = \mathbf{\Delta}$, temos ainda mais $\frac{m(m+1)}{2} - m = \frac{m(m-1)}{2}$ restrições (note que $\mathbf{L}'\mathbf{\Psi}^{-1}\mathbf{L}$, que é simétrica, tem $m(m+1)/2$ parâmetros, os quais devem ser iguais à m parâmetros (em $\mathbf{\Delta}$))
- Assim, $r = \frac{p(p+1)}{2} - \left[p(m+1) - \frac{m(m-1)}{2}\right] = \frac{1}{2}[(p-m)^2 - p - m]$.

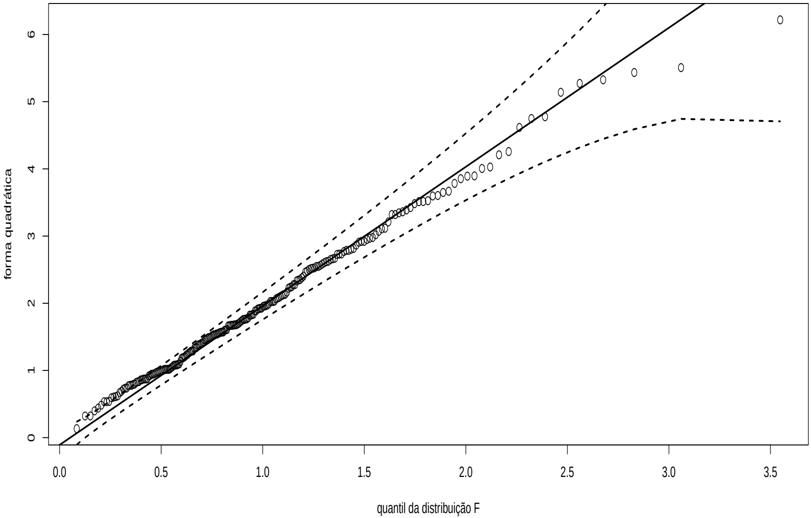
MMV (método de máxima verossimilhança) × MCP (método das componentes principais)

MMV	**MCP**
Inferência através de resultados assintóticos ou de reamostragem	Inferência através de reamostragem
Teste de hipótese para determinar a quantidade de fatores	Determinação de fatores através de metodologias descritivas
Complicado do ponto de vista computacional	Simples do ponto de vista computacional
Há problemas de falta de identificabilidade	Não há problemas de falta de identificabilidade


Gaélico
Inglês
História
Aritmética
Álgebra
Geometria


12
10
8
6
Gaélico
Inglês
História
Aritmética
Álgebra
Geometria


Gaélico
Inglês
História
Aritmética
Álgebra
Geometria
quantil da N(0,1)


forma quadrática
0
1
2
3
4
5
6
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
quantil da distribuição F

- Estimativas das cargas fatoriais: componentes principais

Var.	**Fator 1**	**Fator 2**	**Comun.** $(\sum_{j=1}^{m} l_{ij}^2)$	**Var. Específ.** (ψ_i)
Gaélico	0,646	0,526	0,694	0,306
Inglês	0,738	0,098	0,554	0,446
História	0,447	0,750	0,763	0,237
Artimética	0,772	-0,361	0,726	0,274
Álgebra	0,776	-0,325	0,707	0,293
Geometria	0,691	-0,313	0,576	0,424

- Proporção da soma das variâncias explicadas:

	Fator 1	**Fator 2**
$\widetilde{\lambda}_i$	2,84	1,18
PVE	47,27	19,71
PVEA	47,27	66,98

- $\mathbf{R} - \widetilde{\mathbf{L}}\widetilde{\mathbf{L}}'$

	Gaélico	Inglês	História	Aritmética	Álgebra	Geometria
Gaélico	0,306	-0,102	-0,217	-0,035	-0,004	0,007
Inglês	-0,102	0,446	-0,113	-0,049	-0,109	-0,131
História	-0,217	-0,113	0,237	0,058	0,044	0,055
Aritmética	-0,035	-0,049	0,058	0,274	-0,091	-0,157
Álgebra	-0,004	-0,109	0,044	-0,091	0,293	-0,135
Geometria	0,007	-0,131	0,055	-0,157	-0,135	0,424

- QMresíduos $=$ 0,128.

- $\mathbf{R} - \widetilde{\mathbf{L}}\widetilde{\mathbf{L}}' - \widetilde{\boldsymbol{\Psi}}$

	Gaélico	Inglês	História	Aritmética	Álgebra	Geometria
Gaélico	0,000	-0,102	-0,217	-0,035	-0,004	0,007
Inglês	-0,102	0,000	-0,113	-0,049	-0,109	-0,131
História	-0,217	-0,113	0,000	0,058	0,044	0,055
Aritmética	-0,035	-0,049	0,058	0,000	-0,091	-0,157
Álgebra	-0,004	-0,109	0,044	-0,091	0,000	-0,135
Geometria	0,007	-0,131	0,055	-0,157	-0,135	0,000

- QMresíduos $=$ 0,073.

- A função "factanal" do R estima as cargas fatoriais apenas pelo método de máxima verossimilhança utilizando sempre a matriz de correlações.
- Sintaxe básica: factanal(x=mx,factors=2,rotation="none") ("mx" é a matriz de dados). Também é possível introduzir diretamente a matriz de covariâncias (ou de correlações) amostrais (covmat = matriz de covariâncias ou matriz de correlações) factanal(covmat=sigma,factors=2,rotation="none").
- Se é utilizado uma matriz de dados em que as variáveis não estão padronizados (variância unitária) a função as padroniza. Se é utilizada a matriz de covariâncias ela é transformada na respectiva matriz de correlações.

- Estimativas das cargas fatoriais: máxima verossimilhança

Var.	**Fator 1**	**Fator 2**	**Comun. ($\sum_{j=1}^{m} l_{ij}^2$)**	**Var. Específ. (ψ_i)**
Gaélico	0,583	0,533	0,625	0,375
Inglês	0,641	0,103	0,422	0,578
História	0,354	0,490	0,365	0,635
Aritmética	0,760	-0,297	0,665	0,335
Álgebra	0,742	-0,218	0,599	0,401
Geometria	0,604	-0,143	0,386	0,614

- Proporção da soma das variâncias explicadas:

	Fator 1	**Fator 2**
$\sum_{i=1}^{p} \widetilde{l}_{ij}^2$	2,37	0,691
PVE	39,50	11,50
PVEA	39,50	51,00

- Teste para dois fatores: $\lambda = 3,04(0,55)$. (Observação, para um fator $\lambda = 60,71(< 0,0001)$.
- $\mathbf{R} - \widetilde{\mathbf{L}}\widetilde{\mathbf{L}}'$

	Gaélico	Inglês	História	Aritmética	Álgebra	Geometria
Gaélico	0,375	-0,003	-0,002	-0,011	0,010	0,012
Inglês	-0,003	0,578	0,013	0,028	-0,023	-0,024
História	-0,002	0,013	0,635	0,009	-0,009	-0,015
Aritmética	-0,011	0,028	0,009	0,335	-0,003	-0,012
Álgebra	0,010	-0,023	-0,009	-0,003	0,401	0,023
Geometria	0,012	-0,024	-0,015	-0,012	0,023	0,614

- QMresíduos = 0,093.

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

	Gaélico	Inglês	História	Aritmética	Álgebra	Geometria
Gaélico	0,000	-0,003	-0,002	-0,011	0,010	0,012
Inglês	-0,003	0,000	0,013	0,028	-0,023	-0,024
História	-0,002	0,013	0,000	0,009	-0,009	-0,015
Aritmética	-0,011	0,028	0,009	0,000	-0,003	-0,012
Álgebra	0,010	-0,023	-0,009	-0,003	0,000	0,023
Geometria	0,012	-0,024	-0,015	-0,012	0,023	0,000

- QMresíduos $=$ 0,011.