

Análise Fatorial: parte 1

Prof. Caio Azevedo

Motivação

- Charles Spearman, Thomson, Thurstone e Burt, buscaram obter uma melhor compreensão para “inteligência”. Este conceito está relacionado à várias variáveis cognitivas ([Análise Fatorial](#)).
- Eles conjecturaram a possibilidade de que os desempenhos (notas, variáveis observáveis) de indivíduos (habilidades) em testes cognitivos, em diferentes áreas, pudessem ser representados por um número menor de variáveis não observáveis (latentes, [link 1](#), [link 2](#)).
- Colocando de uma outra forma: tais variáveis latentes podem ser mensuradas através do desempenho dos indivíduos.

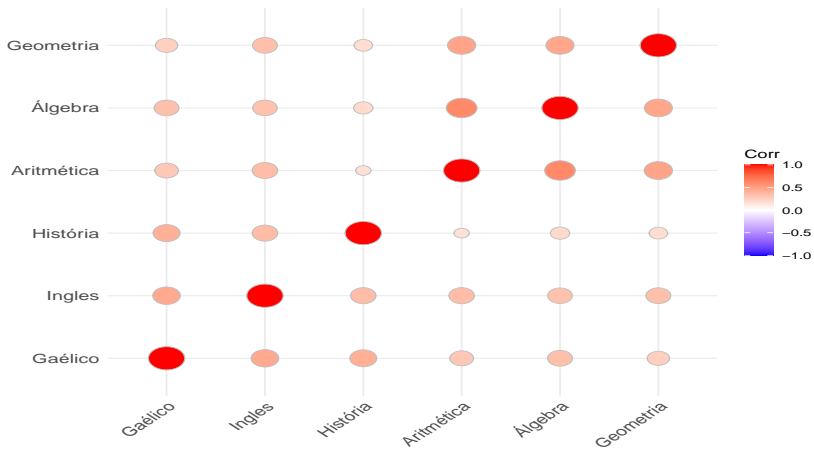
Exemplo 5: Johnson & Wichern (2007)

- 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

Correlações originais



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 $\boldsymbol{\psi} = (1, 0; 1, 5; 0, 8; 1, 3; 1, 7; 2)'$.
- 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 (DS), 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): $\mathbf{X}^* = (\mathbf{X} \times (\text{Chol}(\mathbf{S}^2))^{-1}) \times \text{Chol}(\boldsymbol{\Sigma})$, em que \mathbf{S}^2 é a matriz de covariâncias amostrais de \mathbf{X} .

Modelo de análise fatorial (MAF)

- Seja $\mathbf{X} \sim D_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, em que $\mathcal{E}(\mathbf{X}) = \boldsymbol{\mu}$ e $\text{Cov}(\mathbf{X}) = \boldsymbol{\Sigma}$. Temos que o MAF é dado por:

$$X_1 = \mu_1 + l_{11}F_1 + l_{12}F_2 + \dots + l_{1m}F_m + \xi_1$$

$$X_2 = \mu_2 + l_{21}F_1 + l_{22}F_2 + \dots + l_{2m}F_m + \xi_2$$

$$\vdots = \qquad \qquad \qquad \vdots$$

$$X_p = \mu_p + l_{p1}F_1 + l_{p2}F_2 + \dots + l_{pm}F_m + \xi_p$$

Modelo de análise fatorial (MAF)

- $F_i, i = 1, 2, \dots, m$, são os fatores (ou fatores comuns), $\xi_i, i = 1, \dots, p$ são erros aleatórios (ou fatores específicos) e $l_{ij}, i = 1, \dots, p; j = 1, \dots, m$ são as chamadas cargas fatoriais ($m < p$).
- Várias estruturas podem ser consideradas para F_i e ξ_j . 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, $\mathcal{E}(\mathbf{F}) = \mathbf{0}_m$, $\mathcal{E}(\boldsymbol{\xi}) = \mathbf{0}_p$, $\text{Cov}(\mathbf{F}) = \mathbf{I}_m$, $\text{Cov}(\boldsymbol{\xi}) = \boldsymbol{\Psi}$ e $\text{Cov}(\boldsymbol{\xi}, \mathbf{F}) = \mathbf{0}_{(p \times m)}$, em que

$$\boldsymbol{\Psi} = \begin{bmatrix} \psi_1 & 0 & \dots & 0 \\ 0 & \psi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \psi_p \end{bmatrix}.$$

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 ([link 1](#), [link 2](#)).
- É chamado de ortogonal pois está-se admitindo que os fatores comuns são não correlacionados ($\text{Cov}(\mathbf{F}) = \mathbf{I}_m$).

MAFO: características e propriedades

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

MAFO: características e propriedades

- $Cov(\mathbf{X}) = \Sigma$ (lembrando que $\mathcal{E}(\mathbf{F}) = 0$, $\mathcal{E}(\boldsymbol{\xi}) = 0$), vem que:

$$\begin{aligned}\Sigma &= \mathcal{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'] = \mathcal{E}[(\mathbf{L}\mathbf{F} + \boldsymbol{\xi})(\mathbf{L}\mathbf{F} + \boldsymbol{\xi})'] \\ &= \mathcal{E}(\mathbf{L}\mathbf{F}\mathbf{F}'\mathbf{L}' + \mathbf{L}\mathbf{F}\boldsymbol{\xi}' + \boldsymbol{\xi}\mathbf{F}'\mathbf{L}' + \boldsymbol{\xi}\boldsymbol{\xi}') \\ &= \mathbf{L}\mathcal{E}(\mathbf{F}\mathbf{F}')\mathbf{L}' + \mathbf{L}\mathcal{E}(\mathbf{F}\boldsymbol{\xi}') + \mathcal{E}(\boldsymbol{\xi}\mathbf{F}')\mathbf{L}' + \mathcal{E}(\boldsymbol{\xi}\boldsymbol{\xi}') \\ &= \mathbf{L}Cov(\mathbf{F})\mathbf{L} + \mathbf{L}Cov(\mathbf{F}, \boldsymbol{\xi}) + Cov(\boldsymbol{\xi}, \mathbf{F})\mathbf{L}' + Cov(\boldsymbol{\xi}) = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}\end{aligned}$$

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

MAFO: propriedades

- Note que

$$\mathbf{LL}' = \begin{bmatrix} l_{11} & l_{12} & \cdots & l_{1m} \\ l_{21} & l_{22} & \cdots & l_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ l_{p1} & l_{p2} & \cdots & l_{pm} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & \cdots & l_{p1} \\ l_{12} & l_{22} & \cdots & l_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ l_{1m} & l_{2m} & \cdots & l_{pm} \end{bmatrix} =$$
$$\begin{bmatrix} \sum_{j=1}^m l_{1j}^2 & \sum_{j=1}^m l_{1j}l_{2j} & \cdots & \sum_{j=1}^m l_{1j}l_{pj} \\ \sum_{j=1}^m l_{1j}l_{2j} & \sum_{j=1}^m l_{2j}^2 & \cdots & \sum_{j=1}^m l_{2j}l_{pj} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^m l_{1j}l_{pj} & \sum_{j=1}^m l_{2j}l_{pj} & \cdots & \sum_{j=1}^m l_{pj}^2 \end{bmatrix}$$

MAFO: propriedades

- Variância:

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

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

$$\mathbf{L}^*(\mathbf{L}^*)' + \mathbf{\Psi} = (\mathbf{L}\mathbf{T})(\mathbf{L}\mathbf{T})' + \mathbf{\Psi} = \mathbf{L}\mathbf{T}\mathbf{T}'\mathbf{L}' + \mathbf{\Psi} = \mathbf{L}\mathbf{L}' + \mathbf{\Psi}$$

- Portanto, 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 $\mathbf{Y}|\mathbf{X} = \mathbf{x} \sim N_p(\mathbf{A}\mathbf{x} + \mathbf{B}, \Sigma)$ e $\mathbf{X} \sim N_q(\boldsymbol{\mu}, \Psi)$, então $\mathbf{Y} \sim N_p(\mathbf{A}\boldsymbol{\mu} + \mathbf{B}, \mathbf{A}\Psi\mathbf{A}' + \Sigma)$.
- Exemplo: Se $\mathbf{F} \sim N_m(0, \mathbf{I})$ e $\boldsymbol{\xi} \sim N_p(0, \Psi)$ então $\mathbf{X}|\mathbf{F} = \mathbf{f} \sim N_p(\boldsymbol{\mu} + \mathbf{L}\mathbf{f}, \Psi)$. Assim, $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \mathbf{L}\mathbf{L}' + \Psi)$

MAFO: propriedades

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

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

- Mas $L(\mathbf{L}, \Psi) = L(\mathbf{L}^*, \Psi)$, $\forall \mathbf{L}^* = \mathbf{L}\mathbf{T}$, em que \mathbf{T} é uma matriz ortogonal.
- Falta de identificabilidade: diferentes valores dos parâmetros podem levar ao mesmo valor da verossimilhança.
- Tal aspecto pode ter de ser levado em consideração, consoante o mecanismo de estimação adotado para as cargas fatoriais.

Análise fatorial × Análise de componentes principais

Análise fatorial

Há um modelo estatístico

Impõe uma estrutura específica para matriz de covariâncias dos dados

Busca, por definição, diminuir a dimensionalidade dos dados

Análise de componentes principais

Não há modelo estatístico

Não impõe estrutura

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

$$X_{i1} = \mu_1 + l_{11}F_{i1} + l_{12}F_{i2} + \dots + l_{1m}F_{im} + \xi_{i1}$$

$$X_{i2} = \mu_2 + l_{21}F_{i1} + l_{22}F_{i2} + \dots + l_{2m}F_{im} + \xi_{i2}$$

$$\vdots = \vdots \quad \quad \quad \vdots$$

$$X_{ip} = \mu_p + l_{p1}F_{i1} + l_{p2}F_{i2} + \dots + l_{pm}F_{im} + \xi_{ip}$$

$i=1,2,\dots,n.$

MAFO matricial (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 erros.
- $\mathcal{E}(\mathbf{F}_i) = \mathbf{0}_m$, $\mathcal{E}(\boldsymbol{\xi}_i) = \mathbf{0}_p$, $\text{Cov}(\mathbf{F}_i) = \mathbf{I}_m$, $\text{Cov}(\boldsymbol{\xi}_i) = \boldsymbol{\Psi}$ e $\text{Cov}(\boldsymbol{\xi}_i, \mathbf{F}_j) = \mathbf{0}_{(p \times m)}$, $\forall i, j$, em que $\boldsymbol{\Psi} = \text{diag}(\psi_1, \dots, \psi_p)$, $i, j = 1, 2, \dots, n$.
- Possui uma estrutura semelhante aos MRMNL ([link](#))

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

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

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

Estimação de L : método das componentes principais

- Não requer suposições acerca da distribuição dos erros (fatores específicos (ξ)) nem dos fatores comuns (F).
- Pela **decomposição espectral**, temos que: $\Sigma = \mathbf{E}\Lambda\mathbf{E}'$ (os autovetores ortonormalizados correspondem às colunas da matriz \mathbf{E}), em que

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

Estimação de L : método das componentes principais

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

$$\Sigma = \mathbf{E}\mathbf{\Lambda}\mathbf{E}' = (\mathbf{E}\mathbf{\Lambda}^{1/2})(\mathbf{\Lambda}^{1/2}\mathbf{E}') = \mathbf{L}\mathbf{L}' + \underbrace{\Psi}_0$$
$$\rightarrow \begin{bmatrix} \sqrt{\lambda_1}\mathbf{e}_1 & \sqrt{\lambda_2}\mathbf{e}_2 & \dots & \sqrt{\lambda_p}\mathbf{e}_p \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1}\mathbf{e}'_1 \\ \sqrt{\lambda_2}\mathbf{e}'_2 \\ \dots \\ \sqrt{\lambda_p}\mathbf{e}'_p \end{bmatrix} = \mathbf{L}\mathbf{L}'$$

Estimação de L : método das componentes principais

- Logo

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

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

Estimação de L : método das componentes principais

- Suponha que agora que $m < p$ (m fatores) foram considerados. Dessa forma $\Sigma = \mathbf{E}\mathbf{L}\mathbf{E}' + \mathbf{\Psi}$. Logo (veja equação (1)), uma estimativa para $\psi_i, i = 1, 2, \dots, p$ é dada por

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

em que $\tilde{\sigma}_i^2$ é a variância amostral da variável i .

- Para estimar μ utilizamos o vetor de médias amostrais ($\bar{\mathbf{x}}$), como visto anteriormente.
- Naturalmente desejamos que $\tilde{\Sigma} \approx \tilde{\mathbf{L}}\tilde{\mathbf{L}}' + \tilde{\mathbf{\Psi}}$.
- 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 $\sum_{i=1}^p \tilde{l}_{ij}^2$.
- Pode-se provar que $\sum_{i=1}^p \tilde{l}_{ij}^2 = \tilde{\lambda}_j$.
- É fácil implementar este método no R, usando a função “eigen”, como fora feito.
- Há também, a função “principal” do pacote `psych` (veja arquivo do R no site do curso [aqui](#)).

Observações

- Portanto, 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 $\frac{\tilde{\lambda}_j}{\sum_{i=1}^p \tilde{\sigma}_i^2}$ (PVE pelo fator j).
- Os resultados também se aplicam se trabalharmos com variáveis com variância unitária, ou seja, se usarmos ρ ao invés de Σ .
- Nesse caso (ρ) :

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

Estimação de \mathbf{L} e Ψ : método de máxima verossimilhança

- Maximizar

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

em relação à \mathbf{L} e Ψ .

- Problema: falta de identificabilidade.
- A maximização analítica é complicada. Algum método **númeroico de otimização** tem de ser utilizado (Newton-Raphson, Escore de Fisher, BFGS, Nelder-Mead), veja: [aqui](#).

Formas de se obter identificabilidade

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

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

Cont.

- Note, agora (sendo T uma matriz ortogonal), que:

$$\begin{aligned} L(I + \Delta)^{-1} L' &= LTT'(I + \Delta)^{-1} (LTT')' \\ &= LTT^{-1} (I + \Delta)^{-1} (T^{-1})^{-1} T' L' \\ &= LT(TT^{-1} + T\Delta T^{-1})^{-1} T' L' \\ &= LT(I + T\Delta T')^{-1} T' L' \\ &= L^*(I + T\Delta T')^{-1} (L^*)' \end{aligned}$$

em que $L^* = LT$ e $T\Delta T' \neq \Delta$.

Determinação do número de fatores

- Desejamos que $\tilde{\Sigma} \approx \tilde{\mathbf{L}}\tilde{\mathbf{L}}' + \tilde{\Psi}$.
- 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

$$\frac{\sum_{i=1}^p \tilde{l}_{ij}^2}{\sum_{i=1}^p \tilde{\sigma}_i^2} \text{ (PVE pelo fator } j)$$

- Se trabalharmos com variáveis com variância unitária, ou seja, se usarmos ρ ao invés de Σ , teremos $PVE_i = \frac{\sum_{i=1}^p \tilde{l}_{ij}^2}{p}$
- Testes de hipótese.

Testes de hipótese para determinação do número de fatores via MV

- O resultado é válido se considerarmos Σ ou ρ .
- Desejamos testar se $H_0 : \Sigma = \mathbf{L}\mathbf{L}' + \Psi$ vs $H_1 : \Sigma \neq \mathbf{L}\mathbf{L}' + \Psi$. Seja $\theta = (\mu, \Sigma)$ (irrestrito), $\theta_0 = (\mu, \mathbf{L}, \Psi)$ (restrito).
- Teste da razão de verossimilhanças: Θ_0 - espaço paramétrico sob H_0
 Θ - espaço paramétrico irrestrito.
- Estatística: $\Lambda = \frac{\sup_{\theta \in \Theta_0} L(\mu, \Sigma)}{\sup_{\theta \in \Theta} L(\mu, \Sigma)} = \frac{\sup_{\theta_0} L(\mu, \mathbf{L}, \Psi)}{\sup_{\theta \in \Theta} L(\mu, \Sigma)}$.
- Sob H_0 ou irrestritamente o emv de μ é dado por $\bar{\mathbf{X}}$.

Testes de hipótese para determinação do número de fatores via MV

- Irrestritamente o emv de Σ é dado por

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}}) (\mathbf{x}_i - \bar{\mathbf{x}})'$$

- Os estimadores de MV de \mathbf{L} e Ψ 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 , $\lambda \approx \chi_{(r)}^2$, para n suficientemente grande.
- Vamos determinar r .

Testes de hipótese para determinação do número de fatores via MV

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

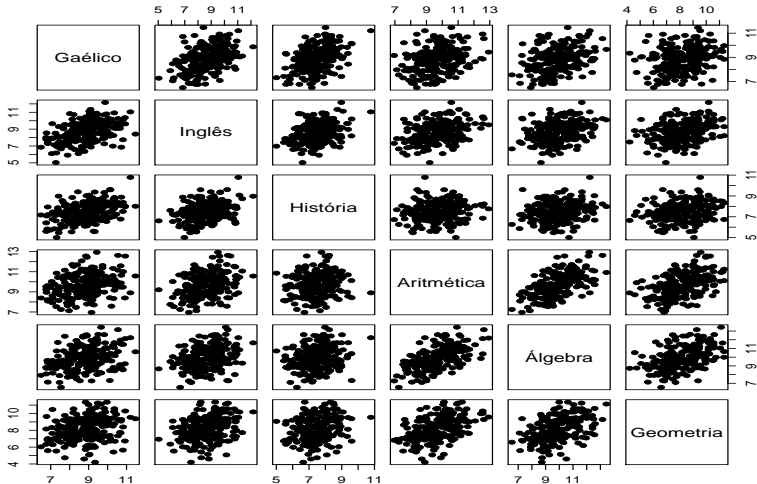
MMV (método de máxima verossimilhança) × MCP (método das componentes principais)

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

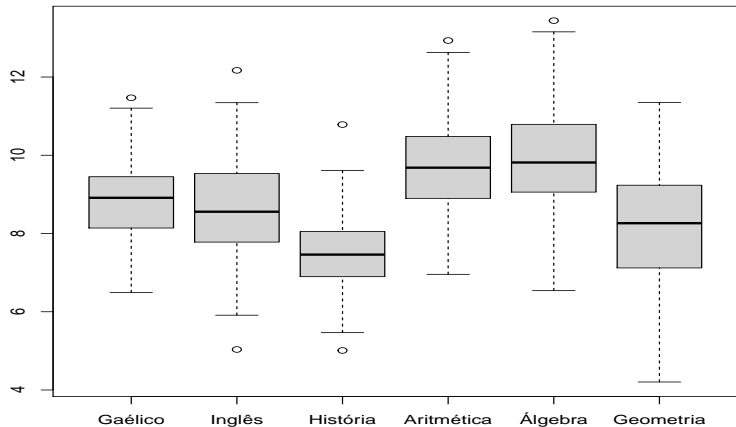
Comentários

- Assim como no caso da [ACP](#), usa-se, na AF, a matriz de correlações.
- Diferentes distribuições podem ser consideradas para os fatores comuns (F) e (ξ), veja por exemplo [aqui](#).
- Com devidas adaptações (e, em geral usando [Teoria Assintótica](#) ou [Métodos de Reamostragem](#)), podemos utilizar ferramentas semelhantes aquelas desenvolvidas para os [MRNLM](#). Veja também ([Análise não assintótica de aproximações para Estatística Multivariada](#)).
- Veja também: [link 1](#), [link 2](#).

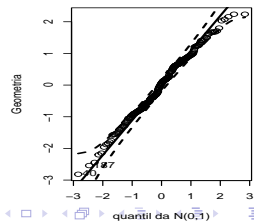
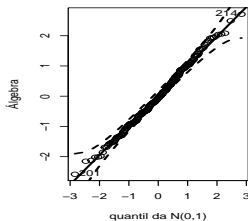
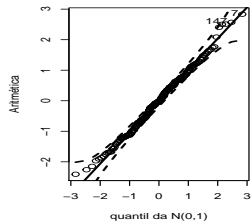
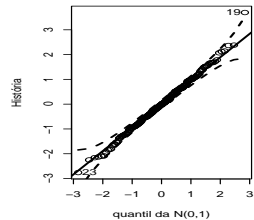
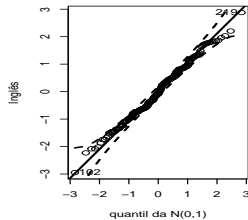
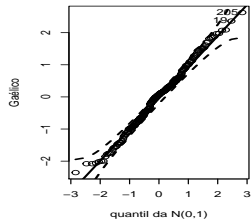
Dispersão entre as variáveis (DS)



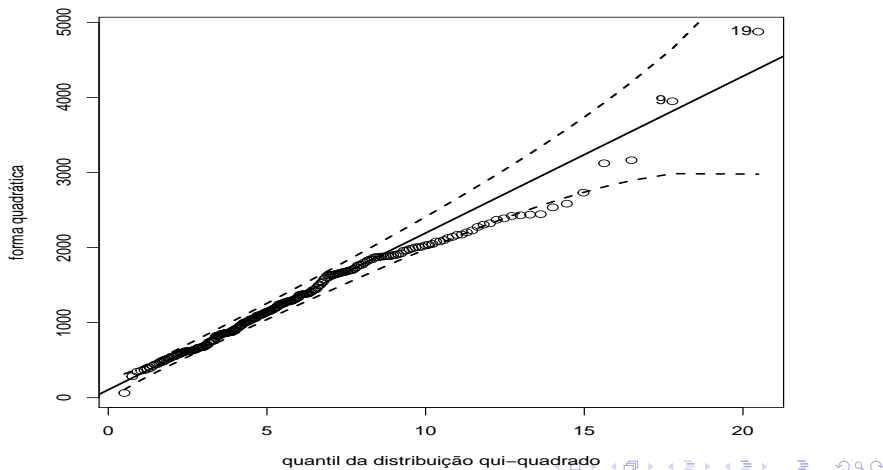
Boxplot das variáveis (DS)



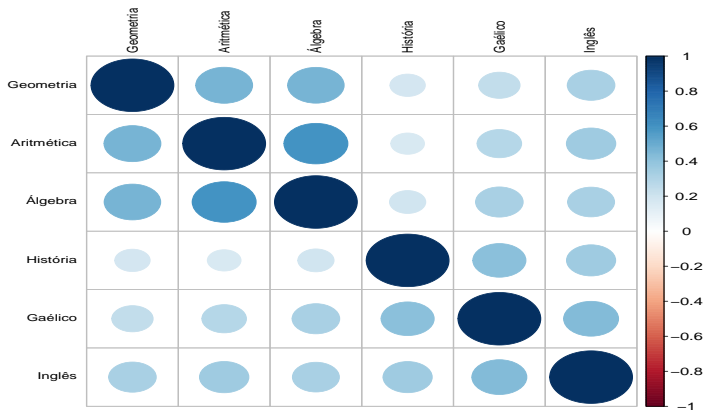
QQplot das variáveis (DS)



QQplot da forma quadrática das variáveis (DS)



Correlações (DS)



■ Estimativas das cargas fatoriais: componentes principais

Var.	Fator 1	Fator 2	Comun. ($\sum_{j=1}^m l_{ij}^2$)	Var. Específ. (ψ_i)
Gaélico	0,658	0,449	0,634	0,366
Inglês	0,688	0,290	0,558	0,442
História	0,517	0,637	0,674	0,326
Aritmética	0,738	-0,413	0,716	0,284
Álgebra	0,744	-0,375	0,694	0,306
Geometria	0,678	-0,355	0,586	0,414

■ Proporção da soma das variâncias explicadas:

	Fator 1	Fator 2
$\tilde{\lambda}_i$	2,73	1,13
PVE	45,55	18,83
PVEA	45,55	64,38

■ $R - \widetilde{LL}'$

	Gaélico	Inglês	História	Aritmética	Álgebra	Geometria
Gaélico	0,366	-0,144	-0,217	-0,012	0,008	-0,039
Inglês	-0,144	0,442	-0,190	-0,034	-0,083	-0,035
História	-0,217	-0,190	0,326	0,045	0,044	0,056
Aritmética	-0,012	-0,034	0,045	0,284	-0,109	-0,177
Álgebra	0,008	-0,083	0,044	-0,109	0,306	-0,174
Geometria	-0,039	-0,035	0,056	-0,177	-0,174	0,414

■ $QM_{\text{resíduos}} = 0,135.$

■ $R - \widetilde{LL}' - \widetilde{\Psi}$

	Gaélico	Inglês	História	Aritmética	Álgebra	Geometria
Gaélico	0,000	-0,144	-0,217	-0,012	0,008	-0,039
Inglês	-0,144	0,000	-0,190	-0,034	-0,083	-0,035
História	-0,217	-0,190	0,000	0,045	0,044	0,056
Aritmética	-0,012	-0,034	0,045	0,000	-0,109	-0,177
Álgebra	0,008	-0,083	0,044	-0,109	0,000	-0,174
Geometria	-0,039	-0,035	0,056	-0,177	-0,174	0,000

■ $QM_{\text{resíduos}} = 0,076.$

- 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), ou seja:

```
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	0,553	0,429	0,490	0,510
Inglês	0,568	0,288	0,406	0,594
História	0,392	0,450	0,356	0,644
Aritmética	0,740	-0,273	0,623	0,377
Álgebra	0,724	-0,211	0,569	0,431
Geometria	0,595	-0,132	0,372	0,628

■ Proporção da soma das variâncias explicadas:

	Fator 1	Fator 2
$\sum_{i=1}^p \tilde{l}_{ij}^2$	2,209	0,606
PVE	36,82	10,09
PVEA	36,82	46,91

- Teste para dois fatores: $\lambda = 2,33(0,670)$. (Observação, para um fator $\lambda = 52(< 0,0001)$).

- $R - \widetilde{LL}'$

	Gaélico	Inglês	História	Aritmética	Álgebra	Geometria
Gaélico	0,510	0,001	0,000	-0,005	0,019	-0,025
Inglês	0,001	0,594	-0,002	0,012	-0,030	0,029
História	0,000	-0,002	0,644	-0,004	0,001	0,007
Aritmética	-0,005	0,012	-0,004	0,377	0,001	-0,007
Álgebra	0,019	-0,030	0,001	0,001	0,431	0,005
Geometria	-0,025	0,029	0,007	-0,007	0,005	0,628

- QMresíduos = 0,097.

■ $R - \widetilde{LL}' - \widetilde{\Psi}$

	Gaélico	Inglês	História	Aritmética	Álgebra	Geometria
Gaélico	-0,000	0,001	0,000	-0,005	0,019	-0,025
Inglês	0,001	0,000	-0,002	0,012	-0,030	0,029
História	0,000	-0,002	0,000	-0,004	0,001	0,007
Aritmética	-0,005	0,012	-0,004	0,000	0,001	-0,007
Álgebra	0,019	-0,030	0,001	0,001	-0,000	0,005
Geometria	-0,025	0,029	0,007	-0,007	0,005	-0,000

■ $QM_{\text{resíduos}} = 0,008.$

Interpretação dos Fatores

- Primeiro fator (média ponderada entre todas as variáveis)
- Segundo fator (“contraste”) entre as variáveis “linguísticas” e as “matemáticas”.
- É possível construir gráficos (incluindo o “Bi-plot”), à semelhança do que é feito para [ACP](#).
- É possível tentar encontrar padrões, interpretações, classificações etc, assim como na [ACP](#).