

# Análise Fatorial

Prof. Caio Azevedo

27 de outubro de 2009

- A partir de uma matriz de dados  $\mathbf{X}_{(n \times p)}$ , calcula-se a matriz  $\mathbf{S}^2$  ou a matriz  $\mathbf{R}$ .
- Componentes principais: calcula-se  $(\tilde{\lambda}_1, \tilde{\mathbf{e}}_1)'$ , ...,  $(\tilde{\lambda}_m, \tilde{\mathbf{e}}_m)'$  através de  $\mathbf{S}^2$  ou  $\mathbf{R}$ . Preferencialmente  $\mathbf{R}$ .
- Teremos  $\tilde{\mathbf{L}}$  e  $\tilde{\Psi}$  a partir de  $\mathbf{S}^2$  e  $\tilde{\mathbf{L}}_Z$  e  $\tilde{\Psi}_Z$  a partir de  $\mathbf{R}$ . Em que:  $\tilde{\mathbf{L}}_Z = \tilde{\mathbf{D}}^{-1} \tilde{\mathbf{L}}$  e  $\tilde{\Psi}_Z = \tilde{\mathbf{D}}^{-1} \tilde{\Psi} \tilde{\mathbf{D}}^{-1}$  e

$$\begin{bmatrix} \tilde{\psi}_1 & 0 & \dots & 0 \\ 0 & \tilde{\psi}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \tilde{\psi}_n \end{bmatrix} = \tilde{\Psi} = \text{diag}(\mathbf{S}^2 - \tilde{\mathbf{L}}\tilde{\mathbf{L}}').$$

- A partir de uma matriz de dados  $\mathbf{X}_{(n \times p)}$ , maximiza-se à verossimilhança em função de  $\boldsymbol{\mu}$ ,  $\mathbf{L}$ ,  $\boldsymbol{\Psi}$ , com as devidas restrições (matriz de covariâncias). Obtem-se assim,  $\tilde{\boldsymbol{\mu}} = \bar{\mathbf{x}}$ ,  $\tilde{\mathbf{L}}$ ,  $\tilde{\boldsymbol{\Psi}}$ ,  $\tilde{\boldsymbol{\Sigma}} = \tilde{\mathbf{L}}\tilde{\mathbf{L}}' + \tilde{\boldsymbol{\Psi}}$  e  $\tilde{\mathbf{D}}$ .
- Posteriormente, calcula-se:  $\tilde{\mathbf{L}}_Z = \tilde{\mathbf{D}}^{-1}\tilde{\mathbf{L}}$ ,  $\tilde{\boldsymbol{\Psi}}_Z = \tilde{\mathbf{D}}^{-1}\tilde{\boldsymbol{\Psi}}\tilde{\mathbf{D}}^{-1}$  e  $\tilde{\mathbf{R}} = \tilde{\mathbf{D}}^{-1}\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{D}}^{-1}$ .

- Considera-se  $\tilde{\mathbf{L}}$ ,  $\tilde{\Psi}$  ou  $\tilde{\mathbf{L}}_Z$ ,  $\tilde{\Psi}_Z$ , como valores conhecidos (componentes principais ou máxima verossimilhança).

- MQP:

- $\tilde{\mathbf{F}}_i = \left( \tilde{\mathbf{L}}' \tilde{\Psi}^{-1} \tilde{\mathbf{L}} \right)^{-1} \tilde{\mathbf{L}}' \tilde{\Psi}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}})$

- $\tilde{\mathbf{F}}_i = \left( \tilde{\mathbf{L}}_Z' \tilde{\Psi}_Z^{-1} \tilde{\mathbf{L}}_Z \right)^{-1} \tilde{\mathbf{L}}_Z' \tilde{\Psi}_Z^{-1} \mathbf{z}_i$

- Regressão:

- $\tilde{\mathbf{F}}_i = \tilde{\mathbf{L}}' \left( \tilde{\mathbf{L}} \tilde{\mathbf{L}}' + \tilde{\Psi}^{-1} \right)^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}).$

- $\tilde{\mathbf{F}}_i = \tilde{\mathbf{L}}' \left( \tilde{\mathbf{L}} \tilde{\mathbf{L}}' + \tilde{\Psi}^{-1} \right)^{-1} \mathbf{z}_i.$

- Notas de indivíduos em diferentes testes. Matriz de correlações baseada em 220 indivíduos. Análise: dados simulados a partir da matriz de correlações.
- Variáveis:
  - Gaelic: idioma gaélico.
  - English: língua inglesa.
  - History: história.
  - Arithmetic: aritmética.
  - Algebra: álgebra.
  - Geometry: geometria.
- Matriz de correlações. Johnson & Wichern.

MAFO:  
componentes  
principais)

MAFO:  
máxima  
verossimil-  
hança)

Escores  
fatoriais)

Exemplo:  
dados sobre  
habilidades  
(notas de  
indivíduos em  
diferentes  
testes)

## ■ Matriz de correlações

	<b>gae.</b>	<b>eng.</b>	<b>hist.</b>	<b>arit.</b>	<b>alg.</b>	<b>geom.</b>
gaelic	1.000	0.439	0.410	0.288	0.329	0.248
english	0.439	1.000	0.351	0.354	0.320	0.329
history	0.410	0.351	1.000	0.164	0.190	0.181
arithmetic	0.288	0.354	0.164	1.000	0.595	0.470
algebra	0.329	0.320	0.190	0.595	1.000	0.464
geometry	0.248	0.329	0.181	0.470	0.464	1.000

MAFO:  
componentes  
principais)MAFO:  
máxima  
verossimil-  
hança)Escores  
fatoriais)Exemplo:  
dados sobre  
habilidades  
(notas de  
indivíduos em  
diferentes  
testes)

## ■ Matriz de correlações

	<b>gae.</b>	<b>eng.</b>	<b>hist.</b>	<b>arit.</b>	<b>alg.</b>	<b>geom.</b>
Gaelic	1.00	0.459	0.477	0.239	0.368	0.213
english	0.459	1.00	0.333	0.371	0.426	0.398
history	0.477	0.333	1.00	0.141	0.198	0.178
arithmetic	0.237	0.371	0.141	1.00	0.558	0.508
algebra	0.368	0.426	0.198	0.558	1.00	0.456
geometry	0.213	0.398	0.178	0.508	0.456	1.00

MAFO:  
componentes  
principais)

MAFO:  
máxima  
verossimil-  
hança)

Escores  
fatoriais)

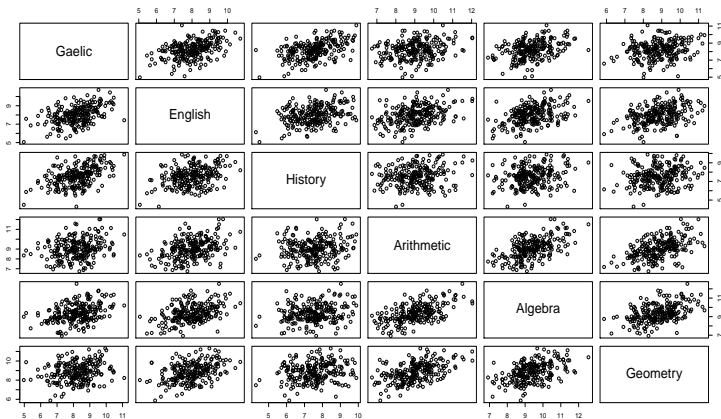
Exemplo:  
dados sobre  
habilidades  
(notas de  
indivíduos em  
diferentes  
testes)

## ■ Medidas descritivas

	<b>gae.</b>	<b>eng.</b>	<b>hist.</b>	<b>arit.</b>	<b>alg.</b>	<b>geom.</b>
Média	8.21	7.99	7.53	9.04	9.38	8.84
DP	1.04	0.99	1.05	1.01	0.973	0.997
CV(%)	12.70	12.34	13.88	11.25	10.35	11.18
Min.	4.97	5.06	4.29	6.74	6.92	5.87
Max.	11.10	10.72	9.89	12.04	12.55	11.32



## ■ Diagrama de dispersão



MAFO: componentes principais)

MAFO: máxima verossimilhança)

Escores fatoriais)

Exemplo: dados sobre habilidades (notas de indivíduos em diferentes testes)

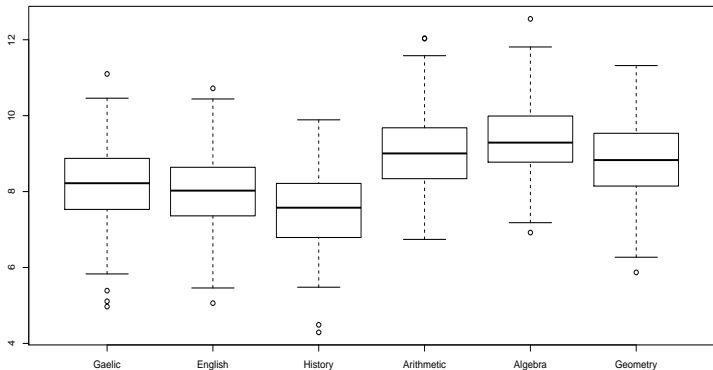
MAFO: componentes principais)

MAFO: máxima verossimilhança)

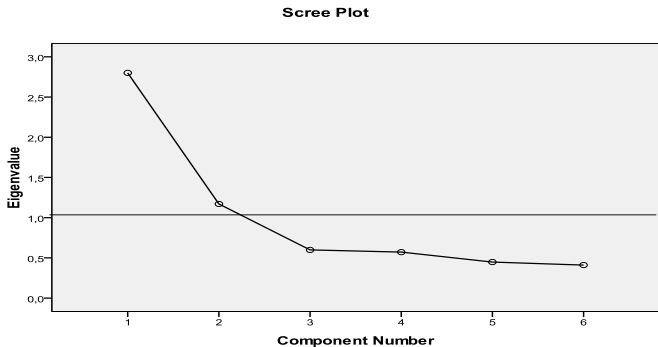
Escores fatoriais)

Exemplo: dados sobre habilidades (notas de indivíduos em diferentes testes)

## ■ Box-plot



## ■ Screeplot (autovalores): Método das componentes principais



- Fatores não rotacionados: componentes principais

<b>Var.</b>	<b>Fator 1</b>	<b>Fator 2</b>	<b>Comun.</b>	<b>Var. Específ.</b>
Gaelic	0.657	0.529	0.711	0.289
English	0.741	0.142	0.569	0.431
History	0.518	0.662	0.706	0.294
Arithmetic	0.709	-0.454	0.710	0.290
Algebra	0.760	-0.269	0.651	0.349
Geometry	0.686	-0.392	0.623	0.377

- Proporção da soma das variâncias explicadas:

	<b>Fator 1</b>	<b>Fator 2</b>
SQ cargas	2.801	1.170
PVE	46.67	19.50
PVEA	46.67	66.18

- Fatores não rotacionados: máxima verossimilhança

<b>Var.</b>	<b>Fator 1</b>	<b>Fator 2</b>	<b>Comun.</b>	<b>Var. Específ.</b>
Gaelic	0.658	0.487	0.670	0.330
English	0.646	0.081	0.424	0.576
History	0.447	0.375	0.340	0.660
Arithmetic	0.659	-0.411	0.603	0.397
Algebra	0.697	-0.216	0.532	0.468
Geometry	0.584	-0.305	0.434	0.566

- Proporção da soma das variâncias explicadas:

	<b>Fator 1</b>	<b>Fator 2</b>
SQ cargas	2.311	0.693
PVE	38.52	11.55
PVEA	38.52	50.07

MAFO: componentes principais)

MAFO: máxima verossimilhança)

Escores fatoriais)

Exemplo: dados sobre habilidades (notas de indivíduos em diferentes testes)

$$\mathbf{R} = \tilde{\mathbf{L}}_Z \tilde{\mathbf{L}}_Z'$$

	<b>gae.</b>	<b>eng.</b>	<b>hist.</b>	<b>arit.</b>	<b>alg.</b>	<b>geom.</b>
Gaelic	0.330	-0.005	0.000	0.004	0.015	-0.023
English	-0.005	0.576	0.014	-0.021	-0.006	0.045
History	0.000	0.014	0.660	0.000	-0.033	0.031
Arithmetic	0.004	-0.021	0.000	0.397	0.010	-0.002
Algebra	0.015	-0.006	-0.033	0.010	0.468	-0.017
Gemoetry	-0.023	0.045	0.031	-0.002	-0.017	0.566

$$\mathbf{Q}M_{\text{resíduos}} = 0.044.$$

$$\mathbf{R} - \tilde{\mathbf{L}}_Z \tilde{\mathbf{L}}_Z' - \tilde{\Psi}_Z$$

MAFO: componentes principais)

MAFO: máxima verossimilhança)

Escores fatoriais)

Exemplo: dados sobre habilidades (notas de indivíduos em diferentes testes)

	gae.	eng.	hist.	arit.	alg.	geom.
Gaelic	0.000	-0.005	0.000	0.004	0.015	-0.023
English	-0.005	0.000	0.014	-0.021	-0.006	0.045
History	0.000	0.014	0.000	0.000	-0.033	0.031
Arithmetic	0.004	-0.021	0.000	0.000	0.010	-0.002
Algebra	0.015	-0.006	-0.033	0.010	0.000	-0.017
Geometry	-0.023	0.045	0.031	-0.002	-0.017	-0.001

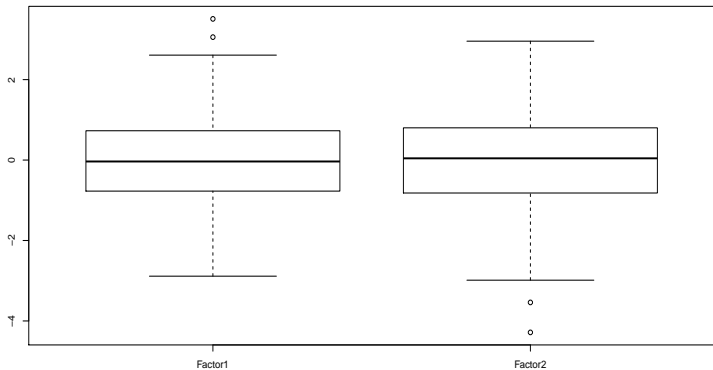
- QMresíduos < 0.001.
- Teste da razão de verossimilhanças  $H_0 : \Sigma = \mathbf{L}\mathbf{L}' + \Psi$ ,  $Q = 4.38$ ,  $gl = 4$  e  $p = 0.3570$ .

- Como interpretar os fatores.
- Solução: rotação (método Varimax).

Var.	Fator 1	Fator 2	Comun.	Var. Específ.
Gaelic	0.192	<b>0.796</b>	0.711	0.289
English	0.444	<b>0.477</b>	0.569	0.431
History	0.102	<b>0.574</b>	0.706	0.294
Arithmetic	<b>0.769</b>	0.108	0.710	0.290
Algebra	<b>0.673</b>	0.282	0.651	0.349
Geometry	<b>0.643</b>	0.141	0.623	0.377



## ■ Box-plot



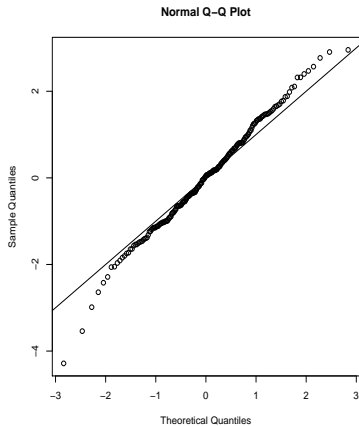
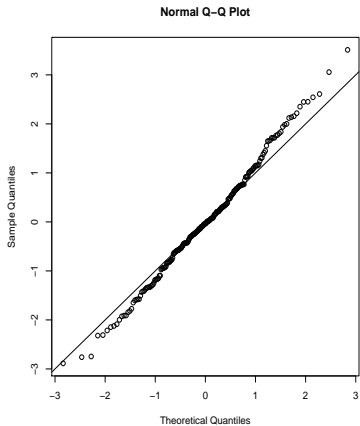
## ■ Box-plot

MAFO: componentes principais)

MAFO: máxima verossimilhança)

Escores fatoriais)

Exemplo: dados sobre habilidades (notas de indivíduos em diferentes testes)



MAFO:  
componentes  
principais)

MAFO:  
máxima  
verossimil-  
hança)

Escores  
fatoriais)

Exemplo:  
dados sobre  
habilidades  
(notas de  
indivíduos em  
diferentes  
testes)

## ■ Box-plot

