

# Análise Multivariada

Lupércio França Bessegato  
Dep. Estatística/UFJF

---

---

---

---

---

---

---

---

## Roteiro

1. Introdução
2. Vetores Aleatórios
3. Normal Multivariada
4. Componentes Principais
5. Análise Fatorial
6. Análise de Agrupamentos
7. Referências

---

---

---

---

---

---

---

---

## Análise Fatorial

---

---

---

---

---

---

---

---

### Exemplo 9.3 – Preferência de Consumidor

- Amostra aleatória de consumidores pontuando atributos de novo produto:
  - √ Respostas em escala semântica de 7 valores
  - √  $X_1$ : Gosto
  - √  $X_2$ : Preço
  - √  $X_3$ : Aroma
  - √  $X_4$ : Adequado para lanche
  - √  $X_5$ : fornece muita energia
  - √ Dados: *BD\_multivariada.xls/preferencia\_consumidor*

---

---

---

---

---

---

---

---

---

---

---

---

### Solução por Componentes Principais:

Variável	Loadings Estimados				Comunalidade		Variância Específica
	$e_{11}$	$e_{12}$	$e_{21}$	$e_{22}$	$h_i^2$	$W_i$	
Gosto	0,331	0,560	0,607	0,816	$(0,560)^2 + (0,816)^2 = 0,979$	$1 - 0,979 = 0,021$	
Preço	0,460	0,777	-0,390	-0,524	$(0,777)^2 + (-0,524)^2 = 0,979$	$1 - 0,979 = 0,021$	
Aroma	0,382	0,645	-0,357	0,788	$(0,645)^2 + (0,788)^2 = 0,979$	$1 - 0,979 = 0,021$	
Adequado lanche	0,556	0,939	-0,078	-0,105	$(0,939)^2 + (-0,105)^2 = 0,983$	$1 - 0,983 = 0,017$	
Energético	0,473	0,728	-0,404	-0,545	$(0,728)^2 + (-0,545)^2 = 0,972$	$1 - 0,972 = 0,028$	
Autovalor	2,853		1,806				
	57,1%		33,2%				

- √ Em geral, uma rotação pode mostrar uma estrutura simples (interpretação simples)

---

---

---

---

---

---

---

---

---

---

---

---

### Análise Fatorial - Minitab

Copy c1-c5 m1  
Print m1

**Data Display**

Matrix M1

```

1,00 0,20 0,96 0,42 0,01
0,20 1,00 0,13 0,71 0,85
0,96 0,13 1,00 0,50 0,11
0,42 0,71 0,50 1,00 0,79
0,01 0,85 0,11 0,79 1,00
    
```

**Stat > Multivariate > Factor Analysis**

Factor Analysis

Variables: C1 X1  
C2 X2  
C3 X3  
C4 X4  
C5 X5

Number of factors to extract: 2

Method of extraction:  Principal components  Maximum likelihood

Type of rotation:  None  Equamax  Varimax  Quartimax  Centmax with gamma

**Factor Analysis - Options**

Matrix to Factor:  Unrotated  Complex  Variables  Covariance  Use matrix: [M]

Loadings for initial solution:  Compute from variables  Use loadings

Rotation method:  Maximum likelihood  Quartimax  Centmax with gamma

Storage:  Loadings: C6-C7  Coefficients:  Scores:  Rotation matrix:  Residual matrix:  H2  Signal matrix:  C8  Eigenvalue matrix:  H3

---

---

---

---

---

---

---

---

---

---

---

---

• Análise Fatorial da Matriz de Correlações:

**Factor Analysis: M1**  
Principal Component Factor Analysis of the Correlation Matrix

Unrotated Factor Loadings and Communalities

Variable	Factor1	Factor2	Communality
Var 1	0,560	0,816	0,979
Var 2	0,777	-0,524	0,879
Var 3	0,645	0,748	0,976
Var 4	0,939	-0,105	0,893
Var 5	0,798	-0,543	0,932
Variance	2,8531	1,8063	4,6594
% Var	0,571	0,361	0,932

$\hat{L}\hat{L}' + \hat{\Psi}$  reproduz aproximadamente R

---

---

---

---

---

---

---

---

---

---

• Cálculo Comunalidades e Resíduos – Minitab:

```
Let C9 = C6*C6 + C7*C7 # Comunalidades
Let C10 = 1 - C9 # Variâncias Específicas
Diagonal C10 M4 # Matriz Ψ
Subtract M4 M2 M5 # Matriz de resíduos
Print M5
```

• Matriz de Resíduos:

**Data Display**  
Matrix M8

0,0000000	0,0126425	-0,0116968	-0,0201455	0,0064418
0,0126425	0,0000000	0,0204813	-0,0749273	-0,0551752
-0,0116968	0,0204813	0,0000000	-0,0275656	0,0011935
-0,0201455	-0,0749273	-0,0275656	0,0000000	-0,0165955
0,0064418	-0,0551752	0,0011935	-0,0165955	0,0000000

---

---

---

---

---

---

---

---

---

---

**Exemplo 9.4 – Ações New York**

• Taxas de retorno de 5 ações negociadas na Bolsa de New York

√ Período: jan/75 a Dez/76  
- Observadas 100 semanas

√ Ações:  
- Allied Chemical  
- du Pont  
- Union Carbide  
- Exxon  
- Texaco

√ Dados: *BD\_multivariada.xls/acoes\_NY*

---

---

---

---

---

---

---

---

---

---



• Comparação soluções:

Variável	Fatores Comuns				Solução			
	F1		F2		1 fator		2 fatores	
	$e_{11}$	$l_{12}$	$e_{21}$	$l_{22}$	$h_1^2$	$\Psi_1$	$h_2^2$	$\Psi_2$
Allied Chemical	0,464	0,783	0,241	0,217	0,615	0,385	0,661	0,339
Du Pont	0,457	0,773	0,308	0,458	0,586	0,414	0,806	0,194
Union Carbide	0,472	0,794	0,261	0,234	0,630	0,370	0,682	0,318
Exxon	0,422	0,713	-0,325	-0,472	0,507	0,493	0,751	0,249
Texaco	0,421	0,712	-0,382	-0,524	0,507	0,493	0,780	0,220
Autovolar	2,056		0,000		57,1%		79,2%	

- ✓ Componentes obtidos a partir de R
- ✓ % acumulada da solução a dois fatores é bem maior que a da solução a um fator

---

---

---

---

---

---

---

---

---

---

• Matriz de Correlações

1,00000	0,57692	0,50866	0,38672	0,46218
0,57692	1,00000	0,59838	0,38952	0,32195
0,50866	0,59838	1,00000	0,43610	0,42563
0,38672	0,38952	0,43610	1,00000	0,52353
0,46218	0,32195	0,42563	0,52353	1,00000

• Matriz de Resíduos:  $R - \hat{LL}' - \hat{\Psi}$

✓  $m = 1$

0	-0,028	-0,114	-0,172	-0,096
-0,028	0	-0,015	-0,161	-0,228
-0,114	-0,015	0	0,130	-0,140
-0,172	-0,161	-0,130	0	0,016
-0,096	-0,228	-0,140	0,016	0

✓  $m = 2$

0	-0,128	-0,164	-0,069	0,018
-0,128	0	-0,123	0,055	0,012
0,164	-0,123	0	0,019	0,017
-0,069	0,055	-0,019	0	-0,231
0,018	0,012	-0,017	-0,231	0

✓ Para  $m=2$   $LL'$  produz números maiores (principalmente  $r_{45}$ )

---

---

---

---

---

---

---

---

---

---

• Interpretação:

Unrotated Factor Loadings and Communalities			
Variable	Factor1	Factor2	Communality
allied_chemical	0,783	0,217	0,661
du_pont	0,773	0,458	0,806
union_carbide	0,794	0,234	0,686
exxon	0,713	-0,472	0,731
texaco	0,712	-0,524	0,781
Variance	2,8565	0,8091	3,6656
% Var	0,571	0,162	0,733

- ✓  $F_1$ : fator de mercado
- ✓ representa condições econômicas gerais
- ✓  $F_2$ : fator industrial
- ✓ contrasta ações de indústrias químicas e de óleo e gás (diferencia setores)
- Em essência, mesma conclusão de ACP (ex. 8.5)

---

---

---

---

---

---

---

---

---

---

### Exemplo 9.5 – Ações New York Solução Máxima Verossimilhança

- Taxas de retorno de 5 ações negociadas na Bolsa de New York

√ Período: jan/75 a Dez/76

– Observadas 100 semanas

√ Ações:

- Allied Chemical
- du Pont
- Union Carbide
- Exxon
- Texaco

√ Dados: *BD\_multivariada.xls/acoes\_NY*

---

---

---

---

---

---

---

---

---

---

---

---

√ Solução por Máxima Verossimilhança

Factor Analysis: allied\_chemical; du\_pont; union\_carbide; exxon; texaco

Maximum Likelihood Factor Analysis of the Correlation Matrix

Unrotated Factor Loadings and Communalities

Variable	Factor1	Factor2	Communality
allied_chemical	0,687	-0,176	0,503
du_pont	0,704	-0,506	0,751
union_carbide	0,685	-0,234	0,525
exxon	0,620	0,086	0,392
texaco	0,782	0,452	0,816
Variance	2,4338	0,5538	2,9876
% Var	0,487	0,111	0,598

√ Solução por Componentes Principais

Principal Component Factor Analysis of the Correlation Matrix

Unrotated Factor Loadings and Communalities

Variable	Factor1	Factor2	Communality
allied_chemical	0,783	0,217	0,661
du_pont	0,773	0,458	0,806
union_carbide	0,794	0,234	0,686
exxon	0,713	-0,472	0,731
texaco	0,712	-0,524	0,781
Variance	2,8565	0,8091	3,6656
% Var	0,571	0,162	0,733

---

---

---

---

---

---

---

---

---

---

---

---

Variável	Máxima Verossimilhança				Componentes Principais			
	$f_{11}$	$f_{12}$	$h^2$	$\psi_1$	$f_{11}$	$f_{12}$	$h^2$	$\psi_1$
Allied Chemical	0,687	-0,176	0,503	0,497	0,783	0,217	0,661	0,339
Du Pont	0,704	-0,506	0,751	0,249	0,773	0,458	0,806	0,194
Union Carbide	0,685	-0,234	0,525	0,475	0,794	0,234	0,685	0,315
Exxon	0,620	0,086	0,392	0,608	0,713	-0,472	0,731	0,269
Texaco	0,782	0,452	0,816	0,184	0,712	-0,524	0,780	0,220
Variância	2,434	0,554	2,988		2,856	0,809	3,663	
% Variância Total	48,7%	11,1%	59,8%		57,1%	16,2%	73,3%	

√ Proporção da variância total amostral padronizada explicada é maior para a fatoração por componentes principais que por máxima verossimilhança

– Componentes principais têm a propriedade de otimizar a variância

√  $F_1$ : loadings positivos e grandes

– não tanto quanto por componentes principais

√  $F_2$ : sinais consistentes com contraste, mas magnitudes em alguns casos são menores

– comparação entre Du Pont e Texaco

---

---

---

---

---

---

---

---

---

---

---

---

• Matriz de Resíduos:  $\hat{R} - \hat{L}\hat{L}' - \hat{\Psi}$

√ Máxima Verossimilhança

m=2					
	0	-0,004	0,004	0,024	-0,004
	-0,004	0	0,003	0,004	0,000
	0,004	0,003	0	-0,031	0,005
	0,024	0,004	-0,031	0	0,000
	-0,004	0,000	0,005	0,000	0

√ Componentes Principais

	0	-0,128	-0,164	-0,069	0,018
	-0,128	0	-0,123	0,055	0,012
	-0,164	-0,123	0	-0,019	-0,017
	-0,069	0,055	-0,019	0	-0,231
	0,018	0,012	-0,017	-0,231	0

√ Elementos da matriz de resíduos são bem menores que aqueles obtidos pela análise fatorial por componentes principais

- Serão escolhidos os resultados da solução por máxima verossimilhança

---

---

---

---

---

---

---

---

---

---

---

---

• Importante:

√ Os padrões dos *loadings* fatoriais iniciais estão restritos pela condição de unicidade da estimativa de  $\hat{L}'\hat{\Psi}\hat{L} = \Delta$

√ Padrões fatoriais úteis frequentemente não são revelados até que os fatores sejam rotacionados

---

---

---

---

---

---

---

---

---

---

---

---

**Exemplo 9.6 – Declato Olímpico**

• Estudo de escores olímpicos de decatlo (Linden, 1977)

√ 160 observações multivariadas (139 atletas)

√ Período: 1948 a 1976

√ Escores padronizados para cada um dos 10 eventos

√ Análise Fatorial de R por componentes principais e por máxima verossimilhança

√ Dados: *BD\_multivariada.xls/decatlo*

---

---

---

---

---

---

---

---

---

---

---

---

√ A distribuição dos scores padronizados são normais ou aproximadamente normais para cada um dos 10 eventos (Linden, 1977)

√ Variáveis:

- X<sub>1</sub>: 100 m rasos
- X<sub>2</sub>: Salto em distância
- X<sub>3</sub>: Arremesso de peso
- X<sub>4</sub>: Salto em altura
- X<sub>5</sub>: 400 m rasos
- X<sub>6</sub>: 100 m com barreiras
- X<sub>7</sub>: Lançamento de disco
- X<sub>8</sub>: Salto com vara
- X<sub>9</sub>: Lançamento de dardo
- X<sub>10</sub>: 1.500 m

---

---

---

---

---

---

---

---

---

---

---

---

• Matriz de Correlações

```

Matrix M1
1,00 0,59 0,35 0,34 0,63 0,40 0,28 0,20 0,11 -0,07
0,59 1,00 0,42 0,51 0,49 0,52 0,31 0,36 0,21 0,09
0,35 0,42 1,00 0,38 0,19 0,36 0,73 0,24 0,44 -0,08
0,34 0,51 0,38 1,00 0,29 0,46 0,27 0,59 0,17 0,18
0,63 0,49 0,19 0,29 1,00 0,34 0,17 0,23 0,13 0,39
0,40 0,52 0,36 0,46 0,34 1,00 0,32 0,33 0,18 0,00
0,28 0,31 0,73 0,27 0,17 0,32 1,00 0,24 0,34 -0,02
0,20 0,36 0,24 0,59 0,23 0,33 0,24 1,00 0,24 0,17
0,11 0,21 0,44 0,17 0,13 0,18 0,34 0,24 1,00 0,00
-0,07 0,09 -0,08 0,18 0,39 0,00 -0,02 0,17 0,00 1,00
    
```

√ Há correlação potencial entre scores sucessivos de atletas que concluíram a prova em mais de uma Olimpíada

- Efetuada análise usando 139 escores representativo de cada atleta
- Escolheu-se aleatoriamente um dos escores dos atletas que participaram de mais de uma Olimpíada

---

---

---

---

---

---

---

---

---

---

---

---

√ Solução por Componentes Principais

```

Factor Analysis: M1
Principal Component Factor Analysis of the Correlation Matrix

Unrotated Factor Loadings and Communalities

Variable Factor1 Factor2 Factor3 Factor4 Communality
Var 1 0,691 -0,217 0,520 -0,206 0,837
Var 2 0,793 -0,184 0,193 0,092 0,701
Var 3 0,702 0,535 -0,047 -0,175 0,811
Var 4 0,674 -0,134 -0,139 0,396 0,640
Var 5 0,620 -0,551 0,084 -0,419 0,870
Var 6 0,687 -0,042 0,161 0,240 0,818
Var 7 0,621 0,521 -0,109 -0,234 0,724
Var 8 0,538 -0,087 -0,411 0,440 0,660
Var 9 0,434 0,438 -0,372 -0,235 0,574
Var 10 0,147 -0,596 -0,658 -0,279 0,888

Variance 3,7866 1,5173 1,1144 0,9134 7,3317
% Var 0,378 0,152 0,111 0,091 0,733
    
```

√ Solução por Máxima Verossimilhança

```

Maximum Likelihood Factor Analysis of the Correlation Matrix

Unrotated Factor Loadings and Communalities

Variable Factor1 Factor2 Factor3 Factor4 Communality
Var 1 0,693 -0,225 0,700 0,170 0,816
Var 2 0,464 -0,281 0,455 -0,240 0,625
Var 3 0,876 -0,234 -0,235 0,040 0,936
Var 4 0,348 -0,407 0,204 -0,437 0,519
Var 5 0,144 -0,563 0,551 0,166 0,669
Var 6 0,418 -0,247 0,307 -0,345 0,449
Var 7 0,666 -0,201 -0,182 0,041 0,669
Var 8 0,211 -0,322 0,127 -0,360 0,294
Var 9 0,392 -0,194 -0,138 -0,058 0,214
Var 10 -0,440 -0,864 -0,070 0,016 0,946

Variance 2,3788 1,8296 1,2648 0,5643 6,0374
% Var 0,238 0,183 0,126 0,056 0,604
    
```

---

---

---

---

---

---

---

---

---

---

---

---





• Matriz de Resíduos:  $R - \hat{L}\hat{L}' - \hat{\Psi}$

√ Componentes Principais

Componentes Principais									
0	-0,075	-0,030	-0,001	-0,047	-0,096	-0,027	0,114	0,051	-0,016
-0,075	0	-0,010	-0,056	-0,077	-0,082	-0,041	-0,042	0,042	0,017
-0,030	-0,010	0	0,042	-0,020	-0,032	-0,031	-0,034	-0,158	0,056
-0,001	-0,056	0,042	0	-0,024	-0,122	-0,001	-0,215	-0,022	0,020
-0,047	-0,077	-0,020	-0,024	0	0,022	-0,017	0,067	0,036	-0,091
-0,096	-0,092	-0,032	-0,122	0,022	0	0,014	-0,129	0,041	0,076
-0,027	-0,041	-0,031	-0,001	-0,017	0,014	0	0,009	-0,254	0,062
0,114	-0,042	-0,034	-0,215	0,067	-0,129	0,009	0	-0,005	-0,109
0,051	0,042	-0,158	-0,022	0,036	0,041	-0,254	-0,005	0	-0,112
-0,016	0,017	0,056	0,020	-0,091	0,076	0,062	-0,109	-0,112	0

√ Máxima Verossimilhança

Máxima Verossimilhança									
0	0,001	0,000	0,013	0,017	-0,014	0,004	-0,001	-0,020	-0,001
0,001	0	0,002	-0,004	-0,002	0,009	-0,021	-0,005	0,003	0,000
0,000	0,002	0	0,005	-0,002	-0,004	0,001	-0,008	0,001	0,000
0,013	-0,004	0,005	0	-0,029	0,001	-0,029	0,002	-0,042	0,002
0,017	-0,002	-0,002	-0,029	0	0,029	-0,002	0,008	-0,050	0,003
-0,014	0,009	-0,004	0,001	0,029	0	0,037	-0,001	-0,010	-0,002
0,004	-0,021	0,001	-0,029	-0,002	0,037	0	0,041	-0,003	0,001
-0,001	-0,005	-0,008	0,002	0,008	-0,001	0,041	0	0,092	-0,001
-0,020	0,003	0,001	-0,042	0,050	-0,010	-0,003	0,092	0	-0,004
-0,001	0,000	0,000	0,002	0,003	-0,002	0,001	-0,001	-0,004	0

---

---

---

---

---

---

---

---

---

---

**Rotação dos Fatores**

- A matriz de covariância  $\Sigma$  é reproduzida pelos loadings fatoriais obtidos por transformação ortogonal, da mesma maneira que os loadings iniciais.

√ Matriz de covariâncias estimada

$$\hat{L}\hat{L}' + \hat{\Psi} = \hat{L}T T' \hat{L}' + \hat{\Psi} = \hat{L}^* \hat{L}^{*'} + \hat{\Psi}$$

√  $T T' = T' T = I$

√  $\hat{L}^*$ : matriz de loadings rotacionados

√ A matriz de resíduos permanece a mesma ( $\hat{h}_i^2$  e  $\hat{\Psi}_i$ )

$$S_n - \hat{L}\hat{L}' - \hat{\Psi} = S_n - \hat{L}^* \hat{L}^{*'} - \hat{\Psi}$$

√ Do ponto de vista estatístico é irrelevante obter  $\hat{L}$  ou  $\hat{L}^*$

---

---

---

---

---

---

---

---

---

---

• Comentários:

- √ Com a rotação, busca-se uma estrutura mais simples
  - loadings originais podem não ter fácil interpretação
- √ Ideal: encontrar um padrão de loadings tais que cada variável carregue-se fortemente em um único fator (com loadings moderados nos outros fatores)
- √ Nem sempre é possível obter esta estrutura mais simples

---

---

---

---

---

---

---

---

---

---

### Exemplo 9.8 – Examination Scores

- √ Lawley & Maxwell (1971)
- √ Avaliações de 220 estudantes do sexo masculino
- √  $p = 6$
- √ Dados: *BD\_multivariada.xls/examination*
- √ Matriz de correlações:

Gaélico	1	0,439	0,410	0,288	0,329	0,248
Inglês	0,439	1	0,351	0,354	0,320	0,329
História	0,410	0,351	1	0,164	0,190	0,181
Aritmética	0,288	0,354	0,164	1	0,595	0,470
Álgebra	0,329	0,320	0,190	0,595	1	0,464
Geometria	0,248	0,329	0,181	0,470	0,464	1

---

---

---

---

---

---

---

---

---

---

### Solução por Máxima Verossimilhança:

```

Factor Analysis: M1
Maximum Likelihood Factor Analysis of the Correlation Matrix
Unrotated Factor Loadings and Communalities

Variable  Factor1  Factor2  Communality
Var 1     0,553    0,428    0,490
Var 2     0,568    0,288    0,406
Var 3     0,392    0,450    0,356
Var 4     0,740    -0,272   0,622
Var 5     0,724    -0,212   0,569
Var 6     0,595    -0,132   0,372

Variance  2,2095   0,6056   2,8151
% Var     0,368   0,101   0,469
    
```

- √  $F_1$ : reflete a resposta global dos estudantes à instrução  
fator de inteligência geral
- √  $F_2$ : não é facilmente identificável
  - Fator “Math – nonmath”
  - metade positiva, metade negativa (fator bipolar)

---

---

---

---

---

---

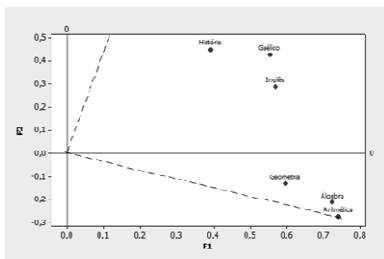
---

---

---

---

### Plot dos loadings fatoriais:



- √ Todos os pontos caem no primeiro quadrante
- √ Revelam-se mais claramente 2 clusters das variáveis

---

---

---

---

---

---

---

---

---

---

- Rotação horária de 20°:  $T = \begin{bmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{bmatrix} = \begin{bmatrix} 0.9397 & 0.3420 \\ -0.3420 & 0.9397 \end{bmatrix}$
- Rotação dos loadings:

Variable	Factor1	Factor2	Communality
Var 1	0,373	0,594	0,490
Var 2	0,435	0,467	0,406
Var 3	0,214	0,558	0,356
Var 4	0,789	0,001	0,622
Var 5	0,752	0,054	0,569
Var 6	0,604	0,083	0,372
Variance	1,9295	0,8856	2,8151
% Var	0,322	0,148	0,469

- √ F<sub>1</sub>\*: variáveis matemáticas do teste com loading alto
  - (desprezíveis em F\*<sub>2</sub>)
  - Fator de habilidade matemática
- √ F<sub>2</sub>\*: variáveis de habilidade verbal com loadings altos
  - Fator de habilidade verbal

---

---

---

---

---

---

---

---

---

---

---

---

• **Comentários:**

- √ O fator de inteligência geral está submergido dos fatores F<sub>1</sub>\* e F<sub>2</sub>\*.
- √ As comunalidades não se modificam (fatores com e sem rotação)

---

---

---

---

---

---

---

---

---

---

---

---

**Exemplo 9.9 – Preferência Consumidor**

- (continuação exemplo 9.3)
  - √ X<sub>1</sub>: Gosto
  - √ X<sub>2</sub>: Preço
  - √ X<sub>3</sub>: Aroma
  - √ X<sub>4</sub>: Adequado para lanche
  - √ X<sub>5</sub>: fornece muita energia
  - √ Dados: *BD\_multivariada.xls/preferencia\_consumidor*

---

---

---

---

---

---

---

---

---

---

---

---





Variável	Loadings Estimados				$h^2$
	Originais		Rotação		
	$\ell_1$	$\ell_2$	$\ell_1$	$\ell_2$	
Allied Chemical	0,687	-0,176	0,600	0,379	0,503
Du Pont	0,704	-0,506	0,851	0,164	0,751
Union Carbide	0,685	-0,234	0,641	0,338	0,525
Exxon	0,620	0,086	0,363	0,510	0,392
Texaco	0,782	0,452	0,208	0,879	0,816
Variância	2,434	0,554	1,670	1,318	2,988
% variabilidade	48,7%	11,1%	33,4%	26,4%	59,8%

- √  $F_1$ \*: Indústrias químicas carregam fortemente
  - Representam condições econômicas que afetam essas ações
- √  $F_2$ \*: Ações de óleo & gás carregam fortemente
  - Representam condições econômicas que afetam essas ações
- √ Rotação tende a destruir um fator geral

---

---

---

---

---

---

---

---

---

---

---

---

### Exemplo 9.11 – Declato Olímpico

- Continuação exemplo 9.6
  - √ 160 observações multivariadas (139 atletas)
  - √ Período: 1948 a 1976
  - √ Escores padronizados para cada um dos 10 eventos
  - √ Análise Fatorial de R por componentes principais e por máxima verossimilhança
  - √ Dados: *BD\_multivariada.xls/decatlo*

---

---

---

---

---

---

---

---

---

---

---

---

- √ A distribuição dos scores padronizados são normais ou aproximadamente normais para cada um dos 10 eventos (Linden, 1977)
- √ Variáveis:
  - $X_1$ : 100 m rasos
  - $X_2$ : Salto em distância
  - $X_3$ : Arremesso de peso
  - $X_4$ : Salto em altura
  - $X_5$ : 400 m rasos
  - $X_6$ : 100 m com barreiras
  - $X_7$ : Lançamento de disco
  - $X_8$ : Salto com vara
  - $X_9$ : Lançamento de dardo
  - $X_{10}$ : 1.500 m

---

---

---

---

---

---

---

---

---

---

---

---

### ✓ Solução por Componentes Principais

**Factor Analysis: MI**  
Principal Component Factor Analysis of the Correlation Matrix

Unrotated Factor Loadings and Communalities

Variable	Factor1	Factor2	Factor3	Factor4	Communality
Var 1	0,493	-0,217	0,700	-0,206	0,837
Var 2	0,789	-0,184	0,193	0,092	0,701
Var 3	0,702	0,281	-0,047	-0,175	0,811
Var 4	0,674	-0,134	-0,139	0,396	0,648
Var 5	0,620	-0,251	0,284	-0,429	0,670
Var 6	0,687	-0,242	0,161	0,345	0,618
Var 7	0,621	0,261	-0,109	-0,234	0,724
Var 8	0,538	-0,287	-0,411	0,440	0,660
Var 9	0,434	0,439	-0,372	-0,235	0,574
Var 10	0,147	-0,396	-0,458	-0,279	0,488
Variance	3,7866	1,5173	1,1144	0,9134	7,3317
% Var	0,378	0,152	0,111	0,091	0,733

### ✓ Solução por Máxima Verossimilhança

Maximum Likelihood Factor Analysis of the Correlation Matrix

Unrotated Factor Loadings and Communalities

Variable	Factor1	Factor2	Factor3	Factor4	Communality
Var 1	0,493	-0,225	0,700	0,179	0,816
Var 2	0,464	-0,381	0,455	-0,240	0,625
Var 3	0,876	-0,194	-0,235	0,040	0,836
Var 4	0,348	-0,407	0,204	-0,437	0,519
Var 5	0,144	-0,563	0,551	0,166	0,669
Var 6	0,418	-0,247	0,307	-0,345	0,449
Var 7	0,666	-0,301	-0,182	0,041	0,569
Var 8	0,211	-0,322	0,127	-0,360	0,294
Var 9	0,392	-0,194	-0,138	-0,058	0,214
Var 10	-0,440	-0,864	-0,070	0,016	0,946
Variance	2,3788	1,8296	1,2648	0,5643	6,0374
% Var	0,238	0,183	0,126	0,056	0,604

### ✓ Rotação Componentes Principais

Rotated Factor Loadings and Communalities  
Varimax Rotation

Variable	Factor1	Factor2	Factor3	Factor4	Communality
Var 1	0,984	0,156	0,156	0,113	0,837
Var 2	0,631	0,194	0,515	0,006	0,701
Var 3	0,245	0,825	0,223	0,148	0,811
Var 4	0,239	0,151	0,750	-0,077	0,648
Var 5	0,797	0,075	0,102	-0,468	0,670
Var 6	0,404	0,153	0,435	0,170	0,648
Var 7	0,186	0,814	0,147	0,079	0,724
Var 8	-0,036	0,116	0,762	-0,217	0,660
Var 9	-0,048	0,725	0,110	-0,141	0,574
Var 10	0,045	-0,041	0,112	-0,934	0,888
Variance	2,1345	2,0230	1,9407	1,2335	7,3317
% Var	0,213	0,202	0,194	0,123	0,733

### ✓ Rotação Máxima Verossimilhança

Maximum Likelihood Factor Analysis of the Correlation Matrix

Unrotated Factor Loadings and Communalities

Variable	Factor1	Factor2	Factor3	Factor4	Communality
Var 1	0,493	-0,225	0,700	0,179	0,816
Var 2	0,464	-0,381	0,455	-0,240	0,625
Var 3	0,876	-0,194	-0,235	0,040	0,836
Var 4	0,348	-0,407	0,204	-0,437	0,519
Var 5	0,144	-0,563	0,551	0,166	0,669
Var 6	0,418	-0,247	0,307	-0,345	0,449
Var 7	0,666	-0,301	-0,182	0,041	0,569
Var 8	0,211	-0,322	0,127	-0,360	0,294
Var 9	0,392	-0,194	-0,138	-0,058	0,214
Var 10	-0,440	-0,864	-0,070	0,016	0,946
Variance	2,3788	1,8296	1,2648	0,5643	6,0374
% Var	0,238	0,183	0,126	0,056	0,604

Variável	Loadings Estimados									
	Componentes Principais				Máxima Verossimilhança					
	$\ell_{11}$	$\ell_{12}$	$\ell_{13}$	$\ell_{14}$	$h^2$	$\ell_{11}$	$\ell_{12}$	$\ell_{13}$	$\ell_{14}$	$h^2$
100m rasos	0,884	0,136	0,156	0,113	0,837	0,166	0,841	0,25022	0,137	0,816
Salto em distância	0,631	0,194	0,515	0,006	0,701	0,232	0,484	0,57987	-0,011	0,625
Arremesso de peso	0,245	0,825	0,223	0,148	0,811	0,927	0,157	-0,2152	0,066	0,936
Salto em altura	0,239	0,151	0,750	-0,077	0,648	0,230	0,168	0,65275	-0,111	0,519
400m rasos	0,797	0,075	0,102	-0,468	0,670	0,058	0,705	0,22856	-0,340	0,669
100m com barreiras	0,404	0,153	0,635	0,170	0,648	0,266	0,270	0,57339	0,069	0,449
Lançamento de disco	0,186	0,814	0,147	0,079	0,724	0,223	0,133	-0,13266	0,009	0,569
Salto com vara	-0,036	0,116	0,762	-0,217	0,660	0,144	0,081	0,50208	-0,121	0,294
Lançamento de dardo	-0,048	0,725	0,110	-0,141	0,574	0,431	0,021	-0,16697	-0,007	0,214
1.500m	0,045	-0,041	0,112	-0,934	0,888	-0,054	0,052	-0,11492	-0,963	0,946
Variância	2,134	2,023	1,941	1,233	7,332	1,771	1,592	1,577	1,097	6,037
% variabilidade	21,3%	20,2%	19,4%	12,3%	73,3%	17,7%	15,9%	15,8%	11,0%	60,4%

- Em ambos os métodos, os loadings fatoriais com rotação apontam para os mesmos atributos subjacentes

✓ Fatores 1 e 2 não estão na mesma ordem



• Solução de máxima verossimilhança -Interpretação:

√ F1\*: Arremesso de peso, lançamento de disco e lançamento de dardo.

- Linden (1977): *explosive arm strength*

√ F2\*: 100 m rasos e 400 m rasos (salto em distância)

- Linden (1977): *running speed*

√ F3\*: Salto em altura, 110 m com barreiras, salto com vara e salto em distância

- Linden (1977): *explosive leg strength*

√ F4\*: 1.500 m e 400 m rasos

- Linden (1977): *running endurance*

---

---

---

---

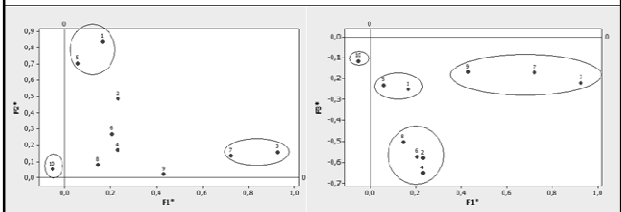
---

---

---

---

• Plot dos *loadings* de máxima verossimilhança com rotação:



√ Em geral, os pontos estão agrupados ao longo dos eixos

---

---

---

---

---

---

---

---

**Exemplo 9.12 – Cálculo dos Scores Fatoriais**

• Ações New York

√ Solução por Máxima Verossimilhança de R

Rotated Factor Loadings

0,599557	0,379038
0,851127	0,164465
0,640928	0,337589
0,363370	0,509725
0,207976	0,879263

Matrix \Psi

0,496861	0,000000	0,000000	0,000000	0,000000
0,000000	0,248534	0,000000	0,000000	0,000000
0,000000	0,000000	0,475245	0,000000	0,000000
0,000000	0,000000	0,000000	0,608143	0,000000
0,000000	0,000000	0,000000	0,000000	0,183642

√ Regressão:  $\hat{f} = \hat{L}^* \mathbf{R}^{-1} z$

√ Mínimos quadrados ponderados:  $\hat{f} = (\hat{L}^* \hat{\Psi}^{-1} \hat{L}^*)^{-1} \hat{L}^* \hat{\Psi}^{-1} z$

---

---

---

---

---

---

---

---

- Para o vetor de observações padronizadas:

$$\sqrt{z'} = [0,5; -1,40; -0,20; -0,70; 1,40]$$

√ scores por regressão

$$\hat{f} = \hat{L}'Rz = \begin{bmatrix} 0,1838 & 0,6626 & 0,2188 & 0,0479 & -0,2085 \\ 0,0402 & -0,1877 & 0,0164 & 0,1123 & 0,8553 \end{bmatrix} \begin{bmatrix} 0,50 \\ -1,40 \\ -0,20 \\ -0,70 \\ 1,40 \end{bmatrix} = \begin{bmatrix} -1,205 \\ 1,398 \end{bmatrix}$$

√ Para cálculo da matriz de dados

$$\hat{f}' = z'R^{-1}\hat{L}'$$

---

---

---

---

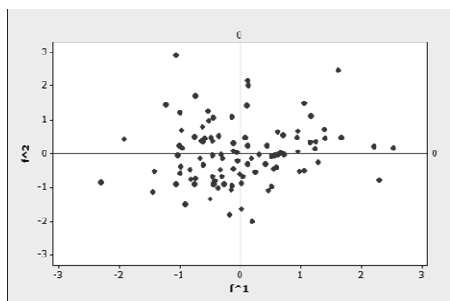
---

---

---

---

- Plot scores fatoriais:




---

---

---

---

---

---

---

---

### Cr1terios para Determina1o do Valor de $m$

- Teste de hip1tese para auxiliar na decis1o do n1mero de fatores ( $m$ ) que s1o suficientes para o modelo de an1lise fatorial
- Suposi1es do teste:
  - √ Os vetores aleat1rios  $Z$  e  $F$  t1m distribu1o normal multivariada
  - √ Amostras de tamanho grande

---

---

---

---

---

---

---

---

• Teste

- √  $H_0$ :  $m$  fatores são suficientes vs.
- $H_1$ : necessários mais que  $m$  fatores

√ Estatística de teste (Bartlett, 1954)

$$T = \left[ n - 1 - \frac{2p + 4m + 5}{6} \right] \ln \left( \frac{|\hat{L}_z \hat{L}_z' + \hat{\Psi}_z|}{|R|} \right)$$

√ Sob  $H_0$ :  $T \sim \chi^2_{gl}$ , com  $gl = \frac{1}{2} [(p - m)^2 - p - m]$

√ Para o teste ser válido  $gl > 0$

- Se  $p = 5, m \leq 2$
- Se  $p = 10, m \leq 5$
- Se  $p = 20, m \leq 14$

√ Para  $n$  grande e  $m$  pequeno em relação a  $p$ , o teste tende a rejeitar  $H_0$  ( indica aumento de  $m$ )

---

---

---

---

---

---

---

---

---

---

• Comentários:

√ Teste somente é válido para dados provenientes de distribuição normal p-variada

√ Para  $n$  grande e  $m$  pequeno em relação a  $p$ , o teste tende a rejeitar  $H_0$  (indicando o aumento de  $m$ )

- Baseando-se apenas na indicação do teste, a tendência será reter no sistema um número muito grande de fatores sem necessidade

---

---

---

---

---

---

---

---

---

---

**Exemplo 9.7 – Teste para valor de  $m$**

• Ações Bolsa New York ( $m = 2$ )

$$\hat{L}_z \hat{L}_z' + \hat{\Psi} = \begin{bmatrix} 1,00000 & 0,57264 & 0,51223 & 0,41107 & 0,45797 \\ 0,57264 & 1,00000 & 0,60103 & 0,39311 & 0,32162 \\ 0,51223 & 0,60103 & 1,00000 & 0,40497 & 0,43013 \\ 0,41107 & 0,39311 & 0,40497 & 1,00000 & 0,52375 \\ 0,45797 & 0,32162 & 0,43013 & 0,52375 & 1,00000 \end{bmatrix}$$

$$R = \begin{bmatrix} 1,00000 & 0,57692 & 0,50866 & 0,38672 & 0,46218 \\ 0,57692 & 1,00000 & 0,59838 & 0,38952 & 0,32195 \\ 0,50866 & 0,59838 & 1,00000 & 0,43610 & 0,42563 \\ 0,38672 & 0,38952 & 0,43610 & 1,00000 & 0,52353 \\ 0,46218 & 0,32195 & 0,42563 & 0,52353 & 1,00000 \end{bmatrix}$$

$$|\hat{L}_z \hat{L}_z' + \hat{\Psi}| = 0,194403 \text{ e } |R| = 0,193234$$

$$T = \left[ 100 - 1 - \frac{2(5) + 4(2) + 5}{6} \right] \ln \left( \frac{0,194403}{0,193234} \right) = 0,574$$

√ Valor crítico:  $gl = \frac{1}{2} [(5 - 2)^2 - 5 - m] = 1$

$\chi^2_1(0,05) = 3,84 \rightarrow$  Não se rejeita  $H_0$

p-valor:  $P\{\chi^2_1 > 0,574\} = 0,448674$

- $H_0$  não deveria ser rejeitada em qualquer nível razoável

---

---

---

---

---

---

---

---

---

---

### Critério de Akaike

Akaike (1974, 1987)

- Suposições:
  - √ Dados provenientes de normal multivariada
  - √ Envolve método de máxima verossimilhança
- Critério: escolhe-se o valor de  $m$  que minimize a função AIC

$$AIC = [-\ln(\max\{L(\mu, \Sigma)\})]$$

- √ Se dois métodos tem a mesma verossimilhança, o procedimento vai privilegiar o modelo com menor número de fatores

---

---

---

---

---

---

---

---

### Critério Bayesiano de Schwarz

- Suposições:
  - √ Dados provenientes de normal multivariada
  - √ Envolve método de máxima verossimilhança
- Critério: escolhe-se o valor de  $m$  que minimize a função SBC

$$SBC = [-\ln(\max\{L(\mu, \Sigma)\})] + \frac{m}{2} \ln(n)$$

- √ Os critérios AIC e SBC devem ser usados com cautela
  - Em geral, indicam quantidade de fatores maior que a necessária
- √ Em geral, o critério SBC resulta em melhores soluções que o método Akaike

---

---

---

---

---

---

---

---

### Matriz de Resíduos

- A observação da matriz de resíduos:
  - √ Muitas vezes, pode indicar quando o número de fatores está superdimensionado
  - √ Ex.:
    - Se  $m$  não for muito pequeno e a matriz de resíduos estiver próxima de zero, recomenda-se testar outras soluções para  $m$  menores que o valor já especificado

---

---

---

---

---

---

---

---

• **Importante:**

√ Análise fatorial deve ser utilizada apenas se utilizada em situações em que as variáveis originais são correlacionadas

√ Consequência:

- Evitar soluções com  $m$  elevado tal que determinados fatores fiquem relacionados com uma única variável original

√ Em situações em que aparecem fatores relacionados a uma única variável  $Z_i$  é recomendável retirar a variável  $Z_i$  e reestimar o modelo de análise fatorial

---

---

---

---

---

---

---

---

**Validação do Modelo**

• Análise Fatorial está fundamentada em suposições que não podem ser verificadas a priori:

√ Linearidade e independência dos fatores

√ Interpretação centrada na informação contida na matriz  $\hat{L}$  (estimada a partir da escolha prévia de  $m$ )

• É importante avaliar até que ponto a matriz  $\hat{L}$  está representando corretamente a relação existente entre as variáveis originais e os fatores do modelo

---

---

---

---

---

---

---

---

**Estratégia para Análise Fatorial**

Johnson & Wichern (2002)

• **Decisões em qualquer Análise Fatorial**

√ Escolha de  $m$ , o número de fatores comuns

- Há muitos testes de adequação assintóticos que são apropriados apenas os dados que são aproximadamente normais

- Os teste provavelmente rejeitarão o modelo para  $m$  pequeno se o número de variáveis e de observações for alto

- Em geral a escolha é baseada em alguma combinação de:

- proporção de variância amostral explicada
- conhecimento do assunto
- razoabilidade dos resultados

---

---

---

---

---

---

---

---

- √ Escolha do método de solução e do tipo de rotação
  - São decisões menos cruciais
  - A maioria de análises fatoriais satisfatórias são aquelas em que:
    - são tentados mais de um método de rotação
    - os resultados confirmam substancialmente a mesma estrutura

---

---

---

---

---

---

---

---

### Roteiro

1. Execute uma Análise Fatorial por componentes principais
  - √ Este método é particularmente apropriado para uma primeira passagem pelos dados
  - √ Procure observações suspeitas plotando os escores fatoriais
    - Calcule os escores padronizados e as distâncias quadráticas para cada observação
  - √ Tente rotação Varimax

---

---

---

---

---

---

---

---

2. Execute Análise Fatorial de Máxima Verossimilhança,
  - √ (incluir uma rotação Varimax)
3. Compare as soluções obtidas pelas duas análises fatoriais
  - √ Os loadings se agrupam da mesma maneira?
  - √ Plote os escores fatoriais obtidos por componentes principais vs. os obtido pela solução de máxima verossimilhança
4. Repita passos anteriores para outros valores de  $m$ 
  - √ Os fatores extras contribuem necessariamente para a compreensão e interpretação dos dados?

---

---

---

---

---

---

---

---

5. Para grandes conjuntos de dados, divida-os pela metade e execute uma Análise Fatorial em cada parte

- ✓ Compare os dois resultados, com aquele obtido do conjunto de dados completo
- ✓ Verifique a estabilidade da solução
- ✓ A divisão pode ser aleatória ou determinística

---

---

---

---

---

---

---

---

### Exemplo 9.14

• Medidas de ossos e crânios de frangos White Leghorn

- ✓ Dados originais: Dunn (1928)
- ✓ Análise fatorial elaborada por Wright (1954)
- ✓ Variáveis:
 

Crânio	Pernas	Asas
✓ $X_1$ : comprimento	✓ $X_3$ : fêmur (comp.)	✓ $X_5$ : úmero (comp.)
✓ $X_2$ : amplitude	✓ $X_4$ : tíbia (comp.)	✓ $X_6$ : cúbito (comp.)
- ✓ Dados: *BD\_multivariada.xls/frangos*

---

---

---

---

---

---

---

---

• Matriz de Correlações amostral:

R (conjunto completo)					
1,000	0,505	0,569	0,602	0,621	0,603
0,505	1,000	0,422	0,467	0,482	0,450
0,569	0,422	1,000	0,926	0,877	0,878
0,602	0,467	0,926	1,000	0,874	0,894
0,621	0,482	0,877	0,874	1,000	0,937
0,603	0,450	0,878	0,894	0,937	1,000

$R_1$  ( $n_1 = 137$ )

1,000	0,696	0,588	0,639	0,694	0,660
0,696	1,000	0,540	0,575	0,606	0,584
0,588	0,540	1,000	0,901	0,844	0,866
0,639	0,575	0,901	1,000	0,835	0,863
0,694	0,606	0,844	0,835	1,000	0,931
0,660	0,584	0,866	0,863	0,931	1,000

$R_2$  ( $n_2 = 139$ )

1,000	0,366	0,572	0,587	0,587	0,598
0,366	1,000	0,352	0,406	0,420	0,386
0,572	0,352	1,000	0,950	0,909	0,894
0,587	0,406	0,950	1,000	0,911	0,927
0,587	0,420	0,909	0,911	1,000	0,940
0,598	0,386	0,894	0,927	0,940	1,000

---

---

---

---

---

---

---

---

• Solução por Componentes Principais

```
Principal Component Factor Analysis of the Correlation Matrix

Unrotated Factor Loadings and Communalities

Variable  Factor1  Factor2  Factor3  Communality
Var 1    0,741  0,250  0,573  0,999
Var 2    0,604  0,721  -0,340  1,000
Var 3    0,929  -0,233  -0,075  0,923
Var 4    0,943  -0,174  -0,067  0,925
Var 5    0,948  -0,143  -0,045  0,920
Var 6    0,945  -0,189  -0,047  0,930
Variance 4,4564  0,7824  0,4584  5,6973
% Var    0,743  0,130  0,076  0,950

Rotated Factor Loadings and Communalities
Varimax Rotation

Variable  Factor1  Factor2  Factor3  Communality
Var 1    0,355  0,902  0,243  0,999
Var 2    0,235  0,211  0,949  1,000
Var 3    0,921  0,218  0,165  0,923
Var 4    0,904  0,251  0,212  0,925
Var 5    0,888  0,284  0,228  0,920
Var 6    0,908  0,264  0,191  0,930
Variance 3,4578  1,1198  1,1197  5,6973
% Var    0,576  0,187  0,187  0,950
```

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

• Solução por Componentes Principais

Variável	Componentes Principais						$h_i^2$	$\Psi_i$
	Loadings Estimados			Loadings Estimados (Rotação)				
	$f_{1i}$	$f_{2i}$	$f_{3i}$	$f_{1i}$	$f_{2i}$	$f_{3i}$		
Comprimento	0,741	0,350	0,573	0,355	0,902	0,243	0,999	0,001
Amplitude	0,604	0,721	-0,340	0,235	0,211	0,949	1,000	0,000
Fêmur	0,929	-0,233	-0,075	0,921	0,218	0,165	0,923	0,077
Tíbia	0,943	-0,174	-0,067	0,904	0,251	0,212	0,925	0,075
Úmero	0,948	-0,143	-0,045	0,888	0,284	0,228	0,920	0,080
Cúbito	0,945	-0,189	-0,047	0,908	0,264	0,191	0,930	0,070
Variância	4,456	0,782	0,458	3,458	1,120	1,120	5,697	
%variabilidade	74,3%	13,0%	7,6%	57,6%	18,7%	18,7%	95,0%	

• Matriz de Resíduos

```
Matrix M8
-0,0000000  0,0002522  0,0055712  0,0028457  -0,0051820  -0,0036794
0,0002522  -0,0000000  0,0031813  -0,0000630  -0,0027691  -0,0006024
0,0055712  0,0031813  -0,0000000  0,0041433  -0,0397322  -0,0468809
0,0028457  -0,0000630  0,0041433  -0,0000000  -0,0478565  -0,0332595
-0,0051820  -0,0027691  -0,0397322  -0,0478565  -0,0000000  0,0127770
-0,0036794  -0,0006024  -0,0468809  -0,0332595  0,0127770  -0,0000000
```

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

• Solução por Máxima Verossimilhança

```
Factor Analysis: M1
Maximum Likelihood Factor Analysis of the Correlation Matrix
* NOTE * Heywood case

Rotated Factor Loadings and Communalities
Varimax Rotation

Variable  Factor1  Factor2  Factor3  Communality
Var 1    0,468  -0,491  0,154  0,484
Var 2    0,209  -0,808  0,066  0,701
Var 3    0,887  -0,281  0,106  0,881
Var 4    0,941  -0,336  -0,026  1,000
Var 5    0,816  -0,349  0,444  0,985
Var 6    0,846  -0,314  0,310  0,910
Variance 3,3169  1,3061  0,3375  4,9605
% Var
```

- ✓ Heywood Case: variância específica da tíbia = 0
- Replicar o resultado, usando a opção Heywood (SAS ou similar)

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---





- Plot dos escores fatoriais obtidos por componentes principais e por máxima verossimilhança
  - √ Se os loadings de um particular fator concordam entre si, os pares de escores deveriam se agrupar próximos à identidade
  - √ Conjuntos de loadings que não concordam produzirão escores fatoriais que se desviam deste padrão
    - Usualmente, associado com o último fator, podendo sugerir que o número de fatores é muito grande (últimos fatores não significativos)

---

---

---

---

---

---

---

---

---

---

√  $F_1$  – Componentes Principais vs. Máxima Verossimilhança

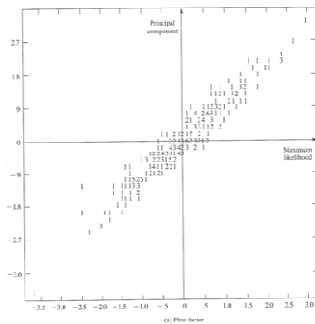


Figure 9.6. Pairs of factor scores for the chicken bone data. (Loadings are estimated by principal component and maximum likelihood methods.)

---

---

---

---

---

---

---

---

---

---

√  $F_2$  – Componentes Principais vs. Máxima Verossimilhança

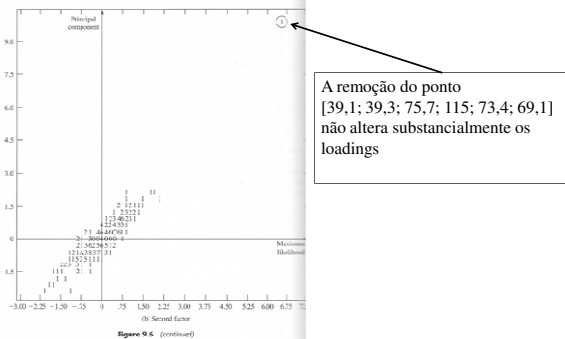


Figure 9.6. (continued)

- √ Uma das observações não é consistente com as demais
  - Tem score incommumente alto

---

---

---

---

---

---

---

---

---

---



• **Conclusões**

- √ Fatores  $F_2^*$  e  $F_3^*$  trocam de posição, mas coletivamente representam as dimensões da cabeça
- √ Fator  $F_1^*$  aparenta ser dimensões do corpo (pernas e asas)
- √ A solução é estável
  - É a mesma interpretação dada pelo conjunto completo
- √ Parece que modelo com um ou dois fatores é suficiente para ajustar os dados

---

---

---

---

---

---

---

---

**Outras Medidas de Ajuste do Modelo**

- √ Critério de Kaiser-Meyer-Olkin (KMO)
- √ Teste de Esfericidade de Bartlett para R

---

---

---

---

---

---

---

---

**Critério de Kaiser-Meyer-Olkin**

- √ Rencher (2002) sugere que para um modelo de Análise Fatorial possa ser ajustado adequadamente aos dados é necessário que  $R^{-1}$  seja próxima da matriz diagonal

- O coeficiente KMO baseia-se nesse princípio

$$KMO = \frac{\sum_{i \neq j} R_{ij}^2}{\sum_{i \neq j} R_{ij}^2 + \sum_{i \neq j} Q_{ij}^2}$$

- √  $R_{ij} = \text{Corr}(X_i, X_j)$
- √  $Q_{ij} =$  Correlação parcial entre 2 variáveis quando todas as outras variáveis são consideradas constantes
- √  $Q_{ij} \approx \text{zero} \rightarrow KMO \approx 1 \rightarrow R^{-1} \approx \text{diagonal}$

---

---

---

---

---

---

---

---

- Adequabilidade do ajuste de um modelo de Análise Fatorial (Rice, 1977)
  - √ modelo adequado:  $KMO \geq 0,8$
  - √ modelo excelente:  $KMO \geq 0,9$
  - √ modelo péssimo:  $KMO \leq 0,5$

---

---

---

---

---

---

---

---

### Teste de Bartlett

- Teste de Esfericidade da matriz de correlação
  - √ Suposições:
    - variáveis provenientes de distribuição normal multivariada
    - modelo de Análise Fatorial pressupõe que as variáveis respostas são correlacionadas entre si
  - √ Teste de hipótese para verificar se a matriz de correlação populacional é próxima ou não da identidade

---

---

---

---

---

---

---

---

- Hipóteses:
  - √  $H_0: \boldsymbol{\rho} = \mathbf{I}$  vs.  $H_1: \boldsymbol{\rho} \neq \mathbf{I}$
- Estatística de teste:

$$T = - \left[ n - \frac{1}{6}(2p + 1) \right] \sum_{i=1}^p \ln \hat{\lambda}_i$$

- √  $\hat{\lambda}_i$ : autovalores da matriz de correlação amostral R
- Sob  $H_0$ :  $T \sim \chi^2_{gl}$ , com  $gl = \frac{1}{2} p(p - 1)$ 
  - √ Para se ajustar o modelo de Análise Fatorial, o teste de Bartlett deve rejeitar  $H_0$ .

---

---

---

---

---

---

---

---

### Comentários

- Análise fatorial permanece muito subjetiva
  - √ Exemplos em que o modelo oferece explicações razoáveis em termos de poucos fatores interpretáveis
  - √ Infelizmente o critério para julgar a qualidade de qualquer análise fatorial não têm sido bem quantificado
- A qualidade do ajuste parece depender do critério Huau (wow)
  - √ Huau, eu compreendi estes fatores

---

---

---

---

---

---

---

---

### Referências

---

---

---

---

---

---

---

---

### Bibliografia Recomendada

- JOHNSON, R. A.; WINCHERN, D. W. *Applied Multivariate Statistical Analysis*. Prentice Hall, 1998
- MINGOTI, D.C. *Análise de Dados através de Métodos de Estatística Multivariada*. Ed. UFMG, 2005.
- LATTIN, J.; CARROLL, J. D.; GREEN, P. E. *Análise de Dados Multivariados*. Cengage, 2011.

---

---

---

---

---

---

---

---