

Análise Fatorial

5

Exemplos

148

Exemplo



- Expectativa de vida
 - ✓ Referente à expectativa de vida (em anos) nos anos 60
 - Médias por país, idade e sexo
 - ✓ Fonte: Keyfitz e Flieger (1971)
 - ✓ Apresentado em Everitt e Hothorn (2011)
 - ✓ Dados: `lifeex.txt`

149

- Variáveis:
 - ✓ Pais: país (fator com 31 níveis)
 - ✓ m0: expectativa média de vida dos homens ao nascer (anos)
 - ✓ m25: expectativa média de vida dos homens na juventude (anos)
 - ✓ m50: expectativa média de vida dos homens na maturidade (anos)
 - ✓ m75: expectativa média de vida dos homens na velhice (anos)
 - ✓ w0: expectativa média de vida das mulheres ao nascer (anos)
 - ✓ w25: expectativa média de vida das mulheres na juventude (anos)
 - ✓ w50: expectativa média de vida das mulheres na maturidade (anos)
 - ✓ w75: expectativa média de vida das mulheres na velhice (anos)



150

- Carregamento dos dados – Comandos em R:

```
> # Carregamento de Dados
> expectativa <- read.table("lifeex.txt", header = T) # Carrega data frame
> colunas <- c("pais", "h0", "h25", "h50", "h75", "m0", "m25", "m50", "m75")
> colnames(expectativa) <- colunas
> head(expectativa)
  pais h0 h25 h50 h75 m0 m25 m50 m75
1  Algeria 63 51 30 13 67 54 34 15
2  Cameroon 34 29 13 5 38 32 17 6
3  Madagascar 38 30 17 7 38 34 20 7
4  Mauritius 59 42 20 6 64 46 25 8
5  Reunion 56 38 18 7 62 46 25 10
6  Seychelles 62 44 24 7 69 50 28 14
> vida <- expectativa[, -1]
```

PPGA
Programa de Pós-Graduação em Administração

Análise de Dados Multivariados - 2017

151

151

- ✓ Teste formal para o número de fatores – Máxima Verossimilhança

```
> # Teste para o número de fatores - MV
>
> rownames(vida) <- expectativa[,1]
> sapply(1:3, function(f)
+ factanal(vida, factors = f, method = "mle")$PVAL)
      objective      objective      objective
1.879555e-24 1.911514e-05 4.578204e-01
```

- ✓ Teste: H_0 : quantidade de fatores é suficiente
- ✓ A solução com 3 fatores pode ser suficiente
- ✓ Cuidado:
 - Há apenas 31 países e o uso de teste assintótico pode ser suspeito

PPGA
Programa de Pós-Graduação em Administração

Análise de Dados Multivariados - 2017

152

152

- ✓ Modelo fatorial com 3 fatores

```
> # Análise fatorial
> vida.af <- factanal(vida, factors = 3, method = "mle")
> vida.af

Call:
factanal(x = vida, factors = 3, method = "mle")

Uniquenesses:
  h0  h25  h50  h75  m0  m25  m50  m75
0.005 0.362 0.066 0.288 0.005 0.011 0.020 0.146

Loadings:
      Factor1 Factor2 Factor3
h0  0.964  0.122  0.226
h25 0.646  0.169  0.438
h50 0.430  0.354  0.790
h75  0.525  0.656
m0  0.970  0.217
m25 0.764  0.556  0.310
m50 0.536  0.729  0.401
m75 0.156  0.867  0.280

By loadings:
      Factor1 Factor2 Factor3
Proportion Var  0.422  0.260  0.205
Cumulative Var  0.422  0.682  0.887

Test of the hypothesis that 3 factors are sufficient.
The chi square statistic is 6.73 on 7 degrees of freedom.
The p-value is 0.458
```

- Default do comando factanal é rotação Varimax

PPGA
Programa de Pós-Graduação em Administração

Análise de Dados Multivariados - 2017

153

153

- ✓ Comunalidades

```
> # Comunalidades:
>
> 1 - vida.af$uniquenesses
      h0      h25      h50      h75      m0      m25      m50      m75
0.9950000 0.6383261 0.9337228 0.7122064 0.9950000 0.9889330 0.9798799 0.8540204
```

- ✓ As variáveis h0, h50, m0, m25 e m50 compartilham muito de suas variâncias através dos fatores
- ✓ As variáveis h25 e h75 são mais únicos

PPGA
Programa de Pós-Graduação em Administração

Análise de Dados Multivariados - 2017

154

154

✓ Interpretação dos fatores

➡

Loadings

> vida.afsloadings

Loadings:

	Factor1	Factor2	Factor3
h0	0.964	0.122	0.226
h25	0.646	0.169	0.438
h50	0.430	0.354	0.790
h75		0.525	0.656
m0	0.970	0.217	
m25	0.764	0.556	0.310
m50	0.536	0.729	0.401
m75	0.156	0.867	0.280

➡

	Factor1	Factor2	Factor3
SS loadings	3.375	2.082	1.640
Proportion Var	0.422	0.260	0.205
Cumulative Var	0.422	0.682	0.887

✓ F1: Expectativa de vida ao nascer (H/M) dominante
(Força de vida ao nascer)

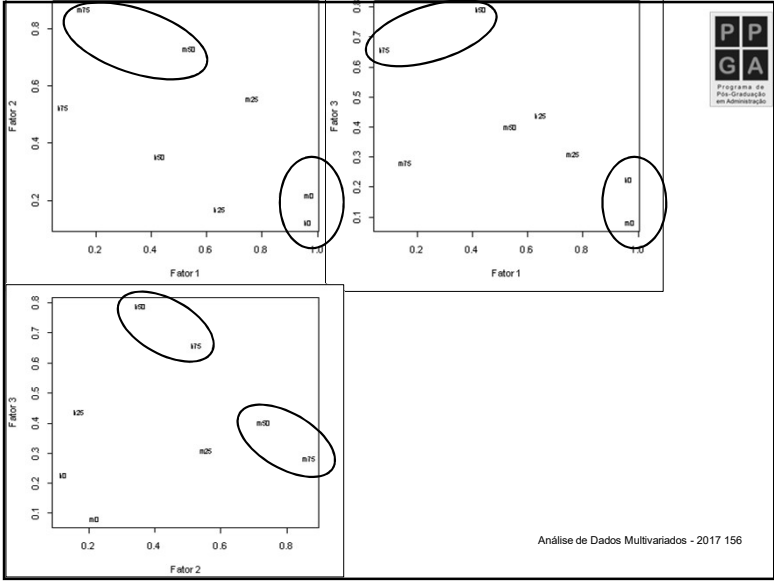
✓ F2: Expectativa de vida em mulheres mais velhas
– (Força de vida mulheres mais velhas)

✓ F3: Cargas maiores para h50 e h75
– (força de vida homens mais velhos)

Análise de Dados Multivariados - 2017

155

155



156

• Gráfico 3-D dos fatores vs. dados

PP
GA
Programa de Pós-Graduação em Administração

3-D Plot dos 3 fatores vs. dados

> library(lattice)

> cloud(Fator3 ~ Fator1 * Fator2, xlim=range(Fator1), ylim=range(Fator2),

+ zlim=range(Fator3),

+ pch=row.names(vida.df),

+ scales = list(distance = rep(1, 3), arrows = FALSE))

>

Análise de Dados Multivariados - 2017

157

157

• Diagramas de dispersão dos escores fatoriais

PP
GA
Programa de Pós-Graduação em Administração

• Força de vida ao nascer variando de países como Camarões e Madagascar para países como USA

• Argélia se sobressai no 3º. Eixo
✓ Alta expectativa de vida entre homens

• Camarões baixa expectativa de vida aos 50

Análise de Dados Multivariados - 2017

158

158

Aplicação

159

Exemplo

- Pesquisa de percepção de marcas:
 - ✓ Avaliação de características relacionadas à marca
 - ✓ Pergunta:
 - Quão [atributo] é a [marca]?
 - ✓ Variáveis:
 - Atributos: *perform, leader, latest, fun, serious, bargain, value, trendy, rebuy*
 - Níveis : 1 (menos) a 10 (mais)
 - brand:
 - Níveis: *a a j*
 - ✓ Respondentes: *100*
 - ✓ Dados: *BD_multivariada.xls/brand*

Análise de Dados Multivariados - 2017

160

160

- Características das marcas – Perguntas:

Atributo	Exemplo de pergunta
<i>perform</i>	Marca tem um forte desempenho?
<i>leader</i>	Marca é líder no mercado?
<i>latest</i>	Marca tem os produtos mais recentes?
<i>fun</i>	Marca é divertida?
<i>serious</i>	Marca é séria?
<i>bargain</i>	Produtos da marca são uma pechincha
<i>value</i>	Produtos da marca possuem um bom valor?
<i>trendy</i>	Marca está na moda?
<i>rebuy</i>	Eu compraria a marca novamente?

- Fonte: Chapman, C.; Feit, E. M. *R for marketing research and analytics*, Springer, 2015

Análise de Dados Multivariados - 2017

161

161

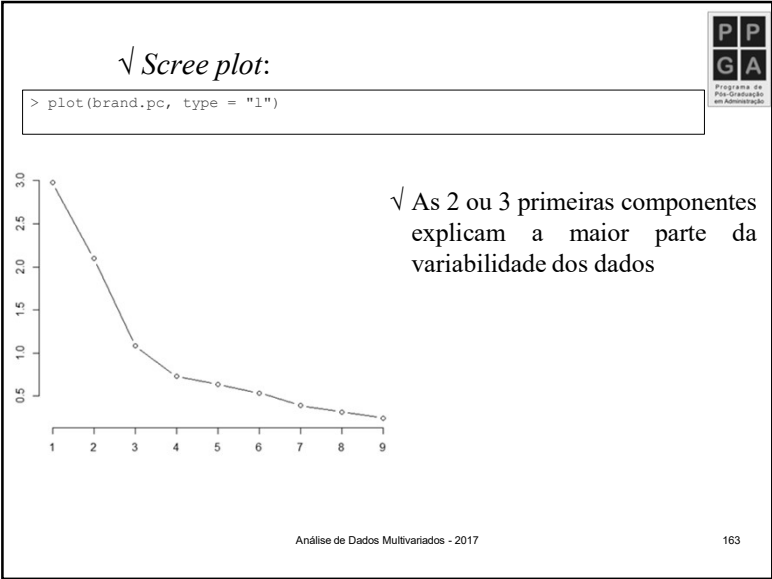
Determinação da Quantidade de Fatores

- *Scree plot*
- Reter fatores associados a autovalores maiores que 1
 - ✓ Quantidade de variância que pode ser atribuída a uma única variável
 - ✓ Fator que captura variância menor que a de uma variável é considerado desprezível

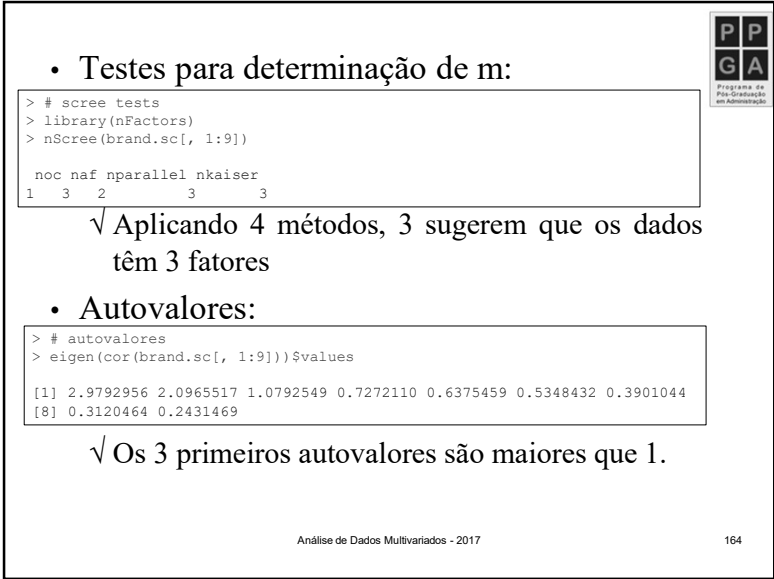
Análise de Dados Multivariados - 2017

162

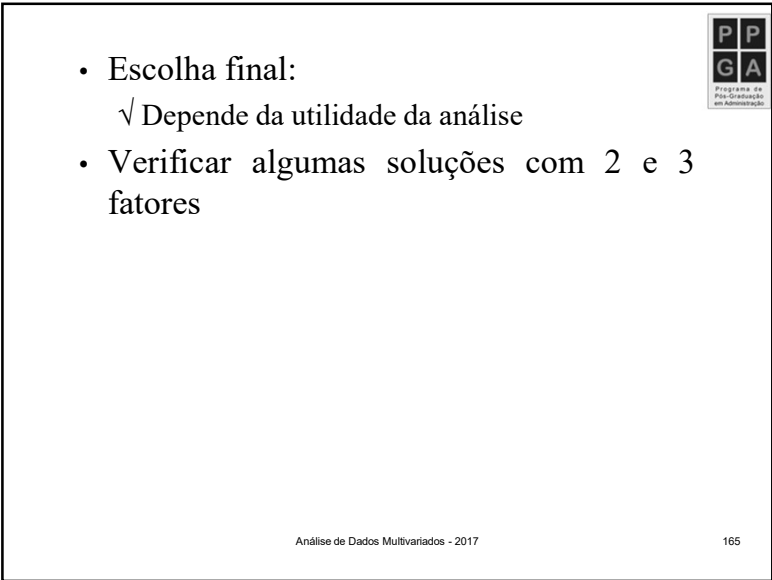
162



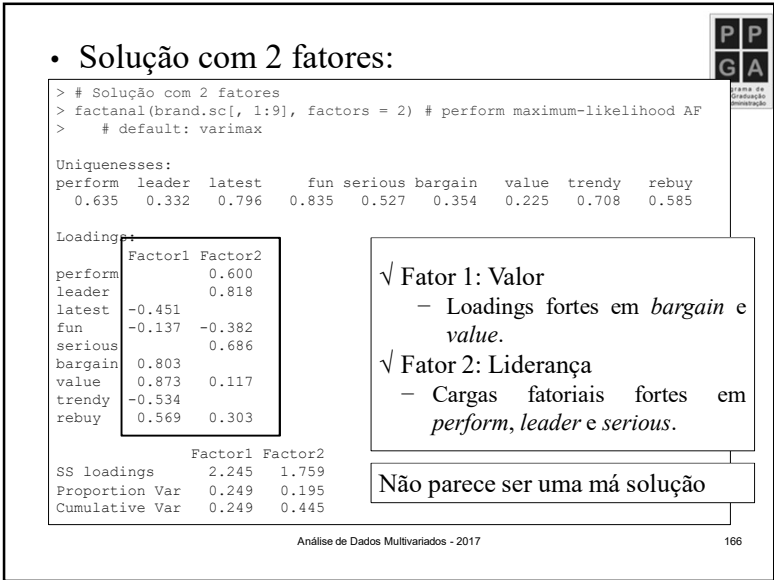
163



164



165



166

PP
GA

Programa de Pós-Graduação em Administração

• Solução com 3 fatores:

```
> # Solução com 3 fatores
> factanal(brand.sc[, 1:9], factors = 3)

Uniquenesses:
perform leader latest fun serious bargain value trendy rebuy
0.624 0.327 0.005 0.794 0.530 0.302 0.202 0.524 0.575

Loadings:
Factor1 Factor2 Factor3
perform 0.607
leader 0.810 0.106
latest -0.163 0.981
fun -0.398 0.205
serious 0.682
bargain 0.826 -0.122
value 0.867 -0.198
trendy -0.356 0.586
rebuy 0.499 0.296 -0.298

SS loadings 1.853 1.752 1.510
Proportion Var 0.206 0.195 0.167
Cumulative Var 0.206 0.401 0.568

Fator 1: Valor
- Cargas fortes em bargain e value.
Fator 2: Liderança no mercado
- Cargas fatoriais fortes em perform, leader e serious.
Fator 3: Atualidade
- Cargas fatoriais fortes em latest e trendy.

Fator adicionado é interpretável
```

Análise de Dados Multivariados - 2017

167

167

PP
GA

Programa de Pós-Graduação em Administração

• Comparação dos modelos:

> # Solução com 2 fatores

Loadings:

Factor1 Factor2

perform 0.600

leader 0.818

latest -0.451

fun -0.137 -0.382

serious 0.686

bargain 0.803

value 0.873 0.117

trendy -0.534

rebuy 0.569 0.303

> # Solução com 3 fatores

Loadings:

Factor1 Factor2 Factor3

perform 0.607

leader 0.810 0.106

latest -0.163 0.981

fun -0.398 0.205

serious 0.682

bargain 0.826 -0.122

value 0.867 -0.198

trendy -0.356 0.586

rebuy 0.499 0.296 -0.298

√ Modelo com 3 fatores:

- Acrescenta na compreensão dos dados conceito claramente interpretável

- Está consistente com sugestões:

- (scree plot, autovalores, scree tests, mapas de percepção)

- Aparece ser superior ao de 2 fatores porque os fatores são melhor interpretáveis

Análise de Dados Multivariados - 2017

168

168

PP
GA

Programa de Pós-Graduação em Administração

Rotação

• Objetivo:

√ Obter novas cargas fatoriais com a mesma proporção de variabilidade

• Tipos:

√ Ortogonal:

- Construtos são independentes

√ Oblíqua:

- Construtos podem estar correlacionados

• Questão:

√ Você deseja permitir que os fatores estejam correlacionados ou não

Análise de Dados Multivariados - 2017

169

169

PP
GA

Programa de Pós-Graduação em Administração

Rotação Oblíqua

• Permitir correlação entre fatores relaciona-se mais com nosso conceito da estrutura latente subjacente e menos com os dados

• Os eixos dimensionais não são perpendiculares, mas assimétricos pelas correlações entre os fatores

Análise de Dados Multivariados - 2017

170

170

PP
GA

Programa de Pós-Graduação em Administração

- No exemplo:
 - ✓ Podemos julgar que os construtos valor e liderança estejam correlacionados
 - ✓ O líder pode colocar um preço especial e, portanto podemos esperar que esses dois construtos sejam correlacionados negativamente (ao invés de independentes)

Análise de Dados Multivariados - 2017171

171

PP
GA

Programa de Pós-Graduação em Administração

- Rotação Oblimin (oblíqua):

```
> library(GPARotation)
> (brand.fa.ob <- factanal(brand.sc[, 1:9], factors = 3, rotation = "oblimin"))
```

Loadings:

	Factor1	Factor2	Factor3
perform	0.601		
leader	0.816		
latest			1.009
fun	-0.381	0.229	
serious		0.689	
bargain	0.859		
value	0.880		
trendy	-0.267	0.128	0.538
rebuy	0.448	0.255	-0.226

Factor Correlations:

	Factor1	Factor2	Factor3
Factor1	1.0000	-0.388	0.0368
Factor2	-0.3884	1.000	-0.1091
Factor3	0.0368	-0.109	1.0000

✓ Não há mudança substancial na interpretação dos fatores

- Loadings ligeiramente diferentes

Resultados apresentam matriz de correlações

Análise de Dados Multivariados - 2017172

172

PP
GA

Programa de Pós-Graduação em Administração

- Varimax e Oblimin – Diferenças:

```
> # Rotação Varimax
```

Loadings:

	Factor1	Factor2	Factor3
perform	0.607		
leader	0.810	0.106	
latest	-0.163		0.981
fun		-0.398	0.205
serious		0.682	
bargain	0.826		-0.122
value	0.867		-0.198
trendy	-0.356		0.586
rebuy	0.499	0.296	-0.298

```
> # Rotação Oblimin
```

Loadings:

	Factor1	Factor2	Factor3
perform	0.601		
leader	0.816		
latest			1.009
fun	-0.381	0.229	
serious		0.689	
bargain	0.859		
value	0.880		
trendy	-0.267	0.128	0.538
rebuy	0.448	0.255	-0.226

- ✓ Mostra separação distinta dos atributos entre os fatores
- ✓ F1 é correlacionado com F2 ($r = -0,39$)
- ✓ Decisão entre as rotações:
 - Basear-se no conhecimento e domínio interpretativo, em vez da estatística

Análise de Dados Multivariados - 2017173

173

PP
GA

Programa de Pós-Graduação em Administração

- Mapa de calor dos loadings:

```
> library(gplots)
> library(RColorBrewer)
> heatmap.2(brand.fa.ob$loadings, col = brewer.pal(9, "Greens"),
+           trace = "none", key = FALSE, dend = "none",
+           Colv = FALSE, cexCol = 1.2,
+           main = "\n\n\nCargas fatoriais para \npercepções de marcas")
```

Cargas fatoriais para percepções de marcas

✓ Separação clara das atributos nos 3 fatores

✓ Rebuy:

- Carrega em F1 (value) e F2(leader)
- Consumidores recompram ou pelo valor da marca ou por ela ter liderança

Análise de Dados Multivariados - 2017174

174

✓ *Path diagram:*

```
> library(semPlot)
> semPaths(brand.fa.ob, what = "est", residuals = FALSE,
+   cut = 0.3, posCol = c("white", "darkgreen"),
+   negCol = c("white", "red"), edge.label.cex = 0.75, nCharNodes = 7)
```

Latentes

✓ Loading +: green
✓ Loading -: red

✓ Ao invés de usar as 9 variáveis observadas, os dados poderiam ser representados com os 3 fatores latentes subjacentes

Variáveis observáveis

Análise de Dados Multivariados - 2017

175

175

- Scores dos fatores para as marcas:
 - ✓ Estimativa da variável latente para cada observação

```
> # Bartlett scores
> brand.fa.ob <- factanal(brand.sc[, 1:9], factors = 3, rotation = "oblimin",
+   scores = "Bartlett")
> brand.scores <- data.frame(brand.fa.ob$scores) # get the factor scores
> brand.scores$brand <- brand.sc$brand # get the matching brands
> head(brand.scores)
```

	Factor1	Factor2	Factor3	brand
1	1.6521364	-0.6886749	0.5256104	a
2	-1.4005333	-1.6681901	-0.6764121	a

✓ Útil em modelos como os de regressão porque pode-se reduzir sua complexidade (número de dimensões)

✓ Permite visualizar os dados em um espaço com quantidade menor de dimensões

Análise de Dados Multivariados - 2017

176

176

- Uso dos escores para Determinar a posição das marcas nos construtos:

```
> # Determinação da posição da marca nos fatores
> brand.fa.mean <- aggregate(. ~ brand, data = brand.scores, mean)
> rownames(brand.fa.mean) <- brand.fa.mean[, 1] # brand names
> brand.fa.mean <- brand.fa.mean[, -1]
> names(brand.fa.mean) <- c("Leader", "Value", "Latest") # factor names
> brand.fa.mean
```

	Leader	Value	Latest
a	0.23158792	-1.06993703	0.39326652
b	0.09686823	1.51913070	0.72391174
c	-0.58937138	1.45069457	-0.07690784
...			

✓ Média de cada marca por construto

Análise de Dados Multivariados - 2017

177

177

✓ Mapa de calor das médias das marcas:

```
> library(gplots)
> library(RColorBrewer)
> heatmap.2(brand.fa.ob$loadings, col = brewer.pal(9, "Greens"),
+   trace = "none", key = FALSE, dend = "none",
+   Colv = FALSE, cexCol = 1.2,
+   main = "\n\n\nCargas fatoriais para \npercepções de marcas")
```

Escore fatorial médio por marca

✓ Média de cada marca por construto

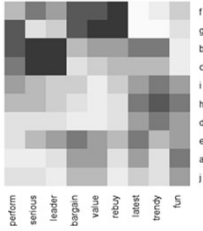
Análise de Dados Multivariados - 2017

178

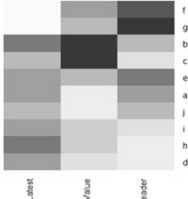
178

• Comparação:

Atributos das Marcas



Escore fatorial médio por marca



✓ Mapa com scores fatoriais é mais simples que a matriz completa das percepções

✓ As similaridades entre as marcas são evidenciadas novamente

– f-g, b-c, ...

Análise de Dados Multivariados - 2017

179

179

Usos da Análise Fatorial

• Examinar a estrutura subjacente e as relações das variáveis

• Reduzir a complexidade dos dados em construtos mais simples e melhor interpretáveis

Análise de Dados Multivariados - 2017

180

180

Referências

Análise de Dados Multivariados - 2017

182

182

Bibliografia Recomendada

• FÁVERO, L. P.; BELFIORE, P; SILVA, F. L.; CHAN, B. L. *Análise de dados: modelagem multivariada para tomada de decisões*. Campus, 2009.

• JOHNSON, R. A.; WINCHERN, D. W. *Applied Multivariate Statistical Analysis*. Prentice Hall, 2007

• LATTIN, J.; CARROLL, J. D.; GREEN, P. E. *Análise de Dados Multivariados*. Cengage Learning, 2005.

• MANLY, B. J. F. *Métodos Estatísticos Multivariados: uma Introdução*. Bookman, 2008.

Análise de Dados Multivariados - 2017

183

183