

2ª LISTA DE EXERCÍCIOS

1. A análise de componentes principais é um dos métodos multivariados mais simples e também um dos mais utilizados. De modo geral, qual(is) o(s) principal(is) objetivo(s) da análise?
2. Defina, em poucas palavras, o que seria um componente principal.
3. A obtenção dos componentes principais é baseada no cálculo de autovalores e autovetores. Numa análise de componentes principais, o que representam os autovalores e os autovetores?
4. A análise de componentes principais é comum codificar (ou padronizar) as variáveis para terem média zero e desvio padrão um. Que justificativa precede esta prática?
5. O que são os escores de um componente principal?
6. A partir de dados de informações nutricionais de sandwiches comercializados por uma rede de fast food, uma análise de componentes principais a partir da matriz de correlações foi utilizada para resumir a informação de sete variáveis nutricionais de dezesseis sandwiches de carne, frango e peixe. Os autovalores e autovetores obtidos foram:

AUTOVALORES

Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7
4.58	1.51	0.39	0.36	0.12	0.03	0.01

AUTOVETORES

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7
Calorias	-0.45	0.14	-0.08	-0.26	0.09	-0.22	0.80
Carboidratos	-0.14	0.74	-0.24	-0.40	-0.32	0.22	-0.26
Gorduras	-0.46	-0.03	0.08	-0.19	0.28	-0.64	-0.51
Gordura Saturada	-0.41	-0.32	0.08	-0.33	0.37	0.68	-0.13
Gorduras Trans	-0.41	-0.26	0.41	0.06	-0.77	0.03	-0.02
Proteínas	-0.38	-0.16	-0.77	0.47	-0.12	0.06	-0.08
Sódio	-0.30	0.49	0.40	0.64	0.27	0.17	0.00

- a) Calcule a porcentagem de variação de cada componente principal.
- b) Nesta análise, apenas dois componentes principais foram utilizados para tomar conclusões acerca dos sandwiches. No seu entendimento, esse é um número suficiente?

Justifique. c) Selecione aquelas variáveis que você julga serem mais importantes para fazer discussões acerca dos dois primeiros componentes. Explique como foi feita a seleção.

d) De acordo com os dois primeiros componentes, quais as principais diferenças entre sandwiches de carne e de frango/peixe?

e) Com base nas informações anteriores, escolha um sandwich para você. Explique quais foram os seus critérios de escolha.

7. Seja X um vetor aleatório $p \times 1$ com $E(X) = \mu$ e $\text{Var}(X) = \Sigma$.

(a) Determine as componentes principais Y_1 e Y_2 para a matriz de covariância

$$\Sigma = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix}.$$

Calcule a proporção da variância total populacional explicada pela primeira componente principal.

(b) Converta a matriz Σ para a matriz de correlação P . Determine as componentes principais Y_1 e Y_2 de P e calcule a proporção total da variância explicada por Y_1 .

8. Obtenha as componentes principais e as proporções de variação total explicadas por cada componente de um vetor aleatório X com variância

$$\Sigma = \begin{bmatrix} \sigma^2 & \sigma^2\rho & 0 \\ \sigma^2\rho & \sigma^2 & \sigma^2\rho \\ 0 & \sigma^2\rho & \sigma^2 \end{bmatrix}.$$

9. O conjunto de dados data (decathlon) disponível na library (FactoMineR) do R contém dados de 41 participações em competições de decatlo, com o nome do atleta na primeira coluna e a competição (Decastar 2004 e Jogos Olímpicos 2004) na última. As provas são apresentadas na seguinte ordem: salto em distância, arremesso de peso, salto em altura, 400 metros rasos, 110 metros com barreiras, lançamento de disco, salto com vara, lançamento de dardo e 1500 metros.

(a) Apresente uma breve análise exploratória dos dados.

(b) Realize uma análise de componentes principais. Justifique a escolha do número de componentes finais e interprete-as.

(c) Proponha um ranking para os atletas.

10. Um naturalista estudou uma população de ursos pardos com o objetivo de manter esta população saudável. Foram observados 61 ursos e obteve-se as seguintes estatísticas:

Variáveis	X_1	X_2	X_3	X_4	X_5	X_6
\bar{x}	95.52	164.38	55.69	93.39	17.98	31.13

Matriz de covariâncias

$$S = \begin{bmatrix} 3266.46 & 1343.97 & 731.54 & 162.68 & 238.37 & & \\ 1343.97 & 721.91 & 324.25 & 537.35 & 80.17 & 117.73 & \\ & 731.54 & 324.25 & 179.28 & 281.17 & 39.15 & 56.80 \\ & 1175.50 & 537.35 & 281.17 & 474.98 & 63.73 & 94.85 \\ & 162.68 & 80.17 & 39.15 & 63.73 & 9.95 & 13.88 \\ & 238.37 & 117.73 & 56.80 & 94.85 & 13.88 & 21.26 \end{bmatrix}.$$

Onde,

X_1 : Peso do urso em kg;

X_2 : Altura do urso em cm;

X_3 : Pescoço do urso em cm;

X_4 : Circunferência do urso em cm;

X_5 : Comprimento da cabeça do urso em cm;

X_6 : Largura da cabeça do urso em cm.

- Faça uma análise de componentes principais a partir da matriz de covariâncias.
- Faça uma análise de componentes principais a partir da matriz de correlação.
- Comente as similaridades e as diferenças entre as duas análises.