

MAE0330 - Análise Multivariada de Dados

2º Semestre/2019

3ª Lista de Exercícios

INFORMAÇÕES IMPORTANTES

- Data de entrega: 11/10 (sexta-feira)
- Forma de entrega: exclusivamente pelo sistema e-Disciplinas (a lista deverá ser entregue no sistema até 23:50 do dia 11/10).
- Podem ser feitas em grupos de **no máximo** 2 alunos.

1. Considere a seguinte matriz de correlação

$$\rho = \begin{pmatrix} 1,000 & -0,488 & 0,150 \\ -0,488 & 1,000 & -0,130 \\ 0,150 & -0,130 & 1,000 \end{pmatrix}.$$

de três variáveis padronizadas Z_1, Z_2 e Z_3 .

- (a) Mostre que ρ pode ser decomposta segundo um modelo fatorial com $m = 1$ fator dado por

$$Z_1 = 0,75F_1 + \epsilon_1,$$

$$Z_2 = -0,65F_1 + \epsilon_2$$

e

$$Z_3 = 0,20F_1 + \epsilon_3,$$

com $\text{Var}(F_1)=1$ e $\text{Cov}(\epsilon_j, F_1)=0$, $j = 1, 2, 3$. Obtenha a matriz Ψ com as variâncias específicas.

- (b) Obtenha as comunalidades e interprete-as.

- (c) Calcule a correlação entre Z_j e F_1 , $j = 1, 2, 3$. Discuta.
2. Considere ainda a matriz de correlação do exercício anterior.
- (a) Obtenha os autovalores e autovetores correspondentes.
- (b) Considerando um modelo fatorial com $m = 1$ fator, obtenha a matriz de cargas fatoriais \mathbf{L} e a matriz de variâncias específicas Ψ usando o método das componentes principais. Compare com os resultados do exercício anterior.
- (c) Obtenha a proporção da variabilidade total dos dados explicada pelo fator.
3. As cargas fatoriais associadas a 6 variáveis padronizadas e as cargas fatoriais rotacionadas (varimax) estão apresentadas a seguir:

Variáveis	Fatores		Fatores Rotac.	
	F_1	F_2	F_1	F_2
Y_1	0,602	0,200	0,484	0,411
Y_2	0,467	0,154	0,375	0,319
Y_3	0,926	0,143	0,603	0,717
Y_4	1,000	0,000	0,519	0,855
Y_5	0,874	0,476	0,861	0,499
Y_6	0,894	0,327	0,744	0,594

- (a) Obtenha as comunalidades e as variâncias específicas para as cargas fatoriais sem e com rotação.
- (b) Qual é proporção da variância total dos dados explicada por cada fator?
- (c) Para uma observação com valores observados das variáveis originais (já padronizados) iguais a (0.8, -0.2, 1.3, -0.6, 1.5, -0.7), obtenha os escores fatoriais utilizando os fatores rotacionados.
4. Os dados no arquivo **T1-9.dat** são referentes a recordes nacionais femininos de corrida para diversos países. As colunas são referentes aos tempos recordes nas seguintes modalidades, respectivamente:
- 100 m (segundos);
 - 200 m (segundos);
 - 400 m (segundos);

- 800 m (minutos);
- 1500 m (minutos);
- 3000 m (minutos); e
- Maratona (minutos).

- Faça uma análise fatorial com a matriz de covariância dos dados.
- Faça uma análise fatorial com a matriz de correlação dos dados.

- Ainda com os dados do arquivo `T1-9.dat`, transforme os tempos recordes em velocidades (na unidade metros por segundo). A maratona corresponde a um percurso de 42.195 metros (ou 26,2 milhas). Faça análise fatorial com a matriz de covariância dos dados e com a matriz de correlação. Discuta os resultados.
- Os vetores $\mathbf{X}^{(1)}$ e $\mathbf{X}^{(2)}$ apresentam os seguintes vetores de média e matriz de covariância:

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{bmatrix} = \begin{bmatrix} -3 \\ 2 \\ 0 \\ 1 \end{bmatrix}$$

e

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} = \begin{bmatrix} 8 & 2 & 3 & 1 \\ 2 & 5 & -1 & 3 \\ 3 & -1 & 6 & -2 \\ 1 & 3 & -2 & 7 \end{bmatrix}$$

- Calcule as correlações canônicas.
 - Obtenha os pares de variáveis canônicas (U_1, V_1) e (U_2, V_2) .
 - Obtenha os autovalores de $\boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21}$ e compare com os autovalores de $\boldsymbol{\Sigma}_{11}^{-1/2} \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1/2}$.
- Quatro diferentes testes foram aplicados em $n = 140$ crianças da sétima série nos Estados Unidos. Os testes aplicados foram:
 - Leitura
 - $X_1^{(1)}$: velocidade
 - $X_2^{(1)}$: capacidade de interpretação

- Matemática
 - $X_1^{(2)}$: velocidade
 - $X_2^{(2)}$: capacidade ou habilidade

A seguinte matriz de correlação foi obtida com os dados:

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{bmatrix} = \left[\begin{array}{cc|cc} 1,0000 & & & \\ 0,6328 & 1,0000 & & \\ \hline 0,2412 & -0,0553 & 1,0000 & \\ 0,0586 & 0,0655 & 0,4248 & 1,0000 \end{array} \right].$$

- Determine as correlações canônicas amostrais.
 - Teste a hipótese $H_0 : \Sigma_{12} = \mathbf{0}$ com nível de significância de 5%. Se a hipótese nula for rejeitada, teste a hipótese da primeira correlação canônica apenas ser igual a zero.
 - Obtenha as variáveis canônicas (utilizando-se os dados padronizados).
 - Obtenha um tabela com as correlações entre as variáveis canônicas e as variáveis originais.
 - Interprete as variáveis canônicas.
8. Os dados disponíveis no arquivo **T7-7.dat** são referentes a propriedades de polpa (ou pasta) de celulose utilizada para fabricação do papel e também algumas propriedades do papel produzido com a polpa. Os dados são de 62 observações e as variáveis observadas são:
- Propriedades do papel
 - $X_1^{(1)}$: BL (*breaking length*);
 - $X_2^{(1)}$: EM (*elastic modulus*);
 - $X_3^{(1)}$: SF (*stress at failure*);
 - $X_4^{(1)}$: BS (*burst strength*).
 - Propriedades da polpa de celulose
 - $X_1^{(2)}$: AFL (*arithmetic fiber length*);
 - $X_2^{(2)}$: LFF (*long fiber fraction*);
 - $X_3^{(2)}$: FFF (*fine fiber fraction*);
 - $X_4^{(2)}$: ZST (*zero span tensile*).

Obtenha os pares de variáveis canônicas e as correlações canônicas. O primeiro par de variáveis canônicas é uma boa medida sumária das variáveis que representam? Justifique a resposta. Teste a significância das correlações canônicas e interprete os pares de variáveis canônicas com correlações significativas com nível de significância igual a 5%.