

ME 731 - Métodos em Análise Multivariada
Segundo semestre de 2020
Lista de Exercícios IV

OBS: Em todos os problemas envolvendo a ACP (Análise de Componentes Principais), além das solicitações constantes em cada uma, utilize as Componentes Principais (CP) para caracterizar as unidades amostrais/experimentais e/ou os grupos, consoante a estrutura de cada conjunto de dados. Além disso, não se esqueça de realizar análises descritivas e inferenciais, conforme visto em sala de aula, sempre que pertinente. Naturalmente, interpretações dos resultados são imprescindíveis.

1. Resolva todos os exercícios deixados em sala.
2. Considere $\mathbf{X} \sim D_2(\mathbf{0}, \mathbf{\Sigma})$, $\mathbf{\Sigma} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, $\rho \in (-1, 1)$. Responda os itens:
 - a) Calcule os autovalores e autovetores (ortonormalizados) associados à $\mathbf{\Sigma}$.
 - b) Obtenha as componentes principais associadas ao vetor \mathbf{X} .
 - c) Interprete as componentes principais que você encontrou no item b).
 - d) O que ocorre com as componentes principais (em termos de suas variâncias e dos autovetores associados à elas) quando $|\rho| \rightarrow 1$. Como as variáveis originais passam a estar relacionadas às componentes principais.
 - e) Repita os itens a), b), c), d) quando $\rho = 0$.
3. Considere $\mathbf{X} \sim D_p(\boldsymbol{\mu}, \mathbf{\Sigma})$, $\mathbf{\Sigma}$ uma matriz positiva definida, com autovalores - autovetores (ortonormalizados) $(\lambda_1, \mathbf{e}'_1)', (\lambda_2, \mathbf{e}'_2)', \dots, (\lambda_p, \mathbf{e}'_p)'$, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$. Prove que a variância generalizada de $\mathbf{\Sigma}$ é igual ao produto dos autovalores de $\mathbf{\Sigma}$. Discuta as vantagens e desvantagens de se utilizar a variância generalizada ao invés da variância total (soma das variâncias) como um critério de escolha do número de componentes principais.
4. Considere o conjunto de dados da Questão 11 da Lista 2. Considerando a matriz de correlações, utilize a metodologia das componentes principais para caracterizar os sexos e os indivíduos, com relação as variáveis medidas, da forma mais completa possível.
5. Para o conjunto de dados anterior, utilize quantas componentes você entender que sejam necessárias, para comparar os sexos, da forma mais completa possível.

6. Com base no conjunto de dados da Questão 12 da Lista II, e considerando a matriz de correlações, utilize a metodologia das componentes principais para caracterizar os atletas, com relação as variáveis medidas, da forma mais completa possível. Faça um rank dos atletas, com base nas componentes principais.
7. Com base no conjunto de dados sobre as tartarugas (Questão 4 da Lista III) e considerando a matriz de correlações, utilize a metodologia das componentes principais para caracterizar os sexos e as tartarugas, com relação as variáveis medidas, da forma mais completa possível.
8. Com relação a Questão anterior, utilize quantas componentes você entender que são necessárias, para comparar os sexos, da forma mais completa possível.
9. Com base no conjunto de dados sobre informações nutricionais de cereais (o arquivo está disponível na página do curso sob o nome de *cereal.txt*), e considerando a matriz de correlações, utilize a metodologia das componentes principais para caracterizar as marcas, em função das variáveis medidas, da forma mais completa possível.
10. Dados relativos à moscas chamadas de “bitting fly” (banco de dados moscas). Os dados se referem à sete variáveis medidas em duas espécies desse tipo de mosca, a saber *Leptoconops carteri* e *Leptoconops torrens* (de agora em diante, *carteri* e *torrens*, por simplicidade). As variáveis (na ordem em que aparecem no banco de dados) são: espécie (0 - *torrens* e 1- *carteri*), comprimento da asa, largura da asa, comprimento do terceiro palpo, largura do terceiro palpo, comprimento do quarto palpo, comprimento do 12º segmento da antena e comprimento do 13º segmento da antena. Considerando a matriz de correlações, utilize a metodologia das componentes principais para caracterizar as espécies e moscas, em função das variáveis medidas, da forma mais completa possível.
11. Com relação a Questão anterior, utilize quantas componentes você entender que são necessárias, para comparar as espécies, da forma mais completa possível.