

IME - PPGEst/UFRGS
Prova II - EST0204 - Estatística Multivariada
Trimestre 2020/03.

Exercício 1. O arquivo “BD1” trás um banco de dados com medidas da mandíbula de 77 cães tailandeses de 5 raças distintas. As nove variáveis estão descritas no arquivo R ”Arquivo prova2 2020”

- (a) Realize a ACP (sem rotação) e apresente os 2 maiores autovalores, os autovetores associados e o percentual acumulado da variação explicada. Interprete a estrutura de correlação dos dados usando os 2 “primeiros” autovetores.
- (b) Ordene os cães baseados nos escores do primeiro CPs e apresente a raça dos 20 com as maiores medidas. Esse ordenamento faz algum sentido? Justifique!
- (c) Realize a AF com 2 fatores (utilizando extração das cargas via máxima verossimilhança e rotação varimax) e apresente as cargas fatoriais, as variâncias dos fatores, o percentual acumulado da variação explicada, as communalidades e a variância não explicada de cada variável. Interprete a estrutura de correlação dos dados usando os 2 primeiros fatores.

Exercício 2. O banco “nacoes2” trás um banco de dados contendo variáveis sócio demográficas de 20 países em um determinado período. Dica: PADRONIZE OS DADOS!!!!!!

- (a) Proceda a uma Análise de Cluster dos países utilizando dois métodos hierárquicos (o complete linkage e o de Ward) e proponha um agrupamento. Descreva os clusters.
- (b) Proceda a uma Análise de Cluster dos países utilizando o método k-means [com centroides iniciais aleatórios e semente set.seed(1)]. Defina um o número k de grupos a priori e justifique a escolha de k. Descreva os clusters.

Exercício 3. Sejam $\mathbf{X}_1 \sim N_3(\mu_1, \Sigma)$ e $\mathbf{X}_2 \sim N_3(\mu_2, \Sigma)$ vetores aleatórios e considere $\mu_1' = [-3, 1, 4]$ e $\mu_2' = [-1, 0, 3]$.

Seja ainda $\Sigma = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$.

- (a) Admitindo os custos de classificação errada $c(2/1) = 100,00$ e $c(1/2) = 50,00$ e probabilidades a priori $p_1 = 0.7$ e $p_2 = 0.3$, encontre e mostre os coeficientes do vetor discriminante e a constante discriminante.
- (b) Considerando $c(2/1) = c(1/2)$ e $p_1 = p_2$, calcule o *TOE* dessa regra. Classifique uma nova observação $x' = [0, 1, 4]$.
- (c) Usando a semente set.seed(1), simule uma amostra de tamanho 100 de cada uma das populações, ajuste a função de Ficher aos dados (usando função lda do R), obtenha a matriz de confusão e o *TAE*. Mostre os coeficientes do vetor discriminante.