

3ª Lista de Exercícios - Análise Multivariada e Aprendizado Não Supervisionado

Exercício 1. Seja $\mathbf{X} \sim N_3(\boldsymbol{\mu}_0, \Sigma_0)$, com $\boldsymbol{\mu} = \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix}$ e $\Sigma = \begin{bmatrix} 1 & 0.8 & -0.4 \\ 0.8 & 4 & -0.56 \\ -0.4 & -0.56 & 2 \end{bmatrix}$.

- (a) Mostre que vetor aleatório \mathbf{X} pode ser representado por; $\mathbf{X} = \boldsymbol{\mu}_0 + \mathbf{L}\mathbf{Z}$, onde $\mathbf{L}\mathbf{L}^\top = \Sigma_0$ (decomposição de Cholesky da Σ_0) e \mathbf{Z} é vetor aleatório de dimensão 3 cujas componentes são independentes com distribuição normal padrão.
- (b) Considerando a representação do vetor aleatório \mathbf{X} em (a), gere uma amostra aleatória $n=30$, verifique se esses dados são gerados da $N_3(\boldsymbol{\mu}_0, \Sigma_0)$
- (c) Dado os dados em (b), construa uma região de 95% de confiança para $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)^\top$ e verifique se μ_0 pertence a região crítica.

Exercício 2. Suponha que tem-se uma amostra aleatória de tamanho $n = 42$ de população $N_2(\boldsymbol{\mu}, \Sigma)$. Dado a amostra observado tem-se o vetor de médias amostrais e as covariâncias resultaram $\mathbf{x} = ((0.564, 0.603)^\top$, e $\mathbf{S} = \begin{pmatrix} 0.0144 & 0.0117 \\ 0.0117 & 0.0146 \end{pmatrix}$

- (a) obtenha a estimativa de máxima verossimilhança de $\boldsymbol{\mu}$ e Σ .
- (b) Obtenha a região 95% de confiança para $\boldsymbol{\mu}$ e faça a representação gráfica do elipsoide de 95% e avalie se $\mu_0 = (0.60, 0.58)$ pertence a região de confiança.
- (c) Avalie os intervalos de 95% de confiança simultâneos para μ_1 e μ_2 .
- (d) Conduza um teste de hipótese para

$$H_0 : \boldsymbol{\mu} = (0.60, 0.58)^\top, \text{ vs } H_1 : \boldsymbol{\mu} \neq (0.60, 0.58)^\top$$

Exercício 3. Considere os dados contidos no arquivo "homeless.csv". Os dados são referentes , de um estudo comparativo entre moradores com moradia e sem moradia, com respeito as variáveis: pontuação físico (PCS, physical component score), pontuação mental (MCS mental component scores) e os sintomas depressivos (DS). No arquivo do conjunto de dados a variável "homeless" é 0 se o indivíduo não é morador de rua e 1 se não tem moradia.

- (a) Avalie se as 3 variáveis para grupo moradores de rua e os que tem moradia pode ser modelados por uma distribuição normal 3-variada.
- (b) Em (a) se os dados não forem normais, transforme os dados considerando a transformação de Box-Cox.
- (c) Formule e teste as hipóteses, para avaliar se em média a pontuação física, mental e sintomas depressivos entre os indivíduos moradores de rua e não moradores de rua são similares. Use $\alpha = 0, 1$.
- (d) Teste a hipótese considerando a região de 90% de confiança.
- (e) Construa intervalos simultâneos de 90% de confiança para diferença de médias entre grupo de moradores de rua e os moradores não moradores.

(f) Refaça o item (e) considerando o método Bonferroni.

Exercício 4. Considere os dados de flores - Iris (originalmente apresentados em Fisher, R. A., 1936)¹, fornece as medidas em centímetros das variáveis: comprimento e largura da sépala e comprimento e largura da pétala, respectivamente, para 50 flores de cada uma das 3 espécies de íris. As espécies são Iris setosa, versicolor e virginica.

- (a) Avalie se há diferenças entre as 3 espécies de flores em quanto as 4 variáveis.
- (b) Supondo que diferenças entre as 3 espécies, construa intervalos de confiança para os possíveis pares de grupos para identificar a origem das diferenças entre as variáveis.

Exercício 5. Resolver exercícios de Johnson and Wichern (2007):

- (a) 5.4, 5.7, 5.9, 5.10 5.20,
- (b) 6.1, 6.2, 6.3, 6.4, 6.7. 6.8.

Referências

- Johnson, R. A. and Wichern, D. W. (2007) Applied Multivariate Statistical Analysis. 5th edition. Prentice-Hall
- Fisher, R. A. (1936) The use of multiple measurements in taxonomic problems. Annals of Eugenics, 7, Part II, 179-188

¹Dados no R `data("iris")`