

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

1. Resolva todos os exercícios deixados em sala.
2. Considere $\mathbf{X}_j \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}), j = 1, \dots, n$, uma amostra aleatória em que $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ são ambos desconhecidos. Obtenha a estatística do teste da razão de verossimilhanças (TRV), e sua respectiva distribuição assintótica sob H_0 , para testar $H_0 : \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$ vs $H_1 : \boldsymbol{\Sigma} \neq \boldsymbol{\Sigma}_0$, em que $\boldsymbol{\Sigma}_{0(p \times p)}$ é uma matriz conhecida.
3. Utilize o resultado da Questão 2 para testar a referida hipótese considerando $(\alpha = 5\%)$, nos seguintes casos:

a) $\boldsymbol{\Sigma}_0 = \begin{bmatrix} 0,03 & 0 \\ 0 & 0,01 \end{bmatrix}$

b) $\boldsymbol{\Sigma}_0 = \begin{bmatrix} 0,02 & 0,01 \\ 0,01 & 0,02 \end{bmatrix}$

Utilize as variáveis *Petal.Length* e *Petal.Width* e o grupo *iris setosa* do conjunto de dados “iris”.

O que significa cada hipótese? Para cada item anterior, apresente a matriz de covariâncias estimada, o valor da estatística do teste, o nível descritivo e a respectiva conclusão.

4. Utilize o TRV para testar a igualdade de matrizes de variâncias-covariâncias entre dois grupos, visto em sala de aula, para testar as referidas hipóteses considerando $(\alpha = 5\%)$, tendo como população 1, o grupo *iris setosa* e a população 2, o grupo *iris versicolor*. Considere as variáveis *Petal.Length* e *Petal.Width*. Apresente as matrizes de covariâncias estimadas de cada

grupo e a ponderada, o valor da estatística do teste, o nível descritivo e a respectiva conclusão.

5. Considere $\mathbf{X}_{ij} \sim N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, $i = 1, 2, j = 1, \dots, n_i$, amostras aleatórias de duas populações independentes, em que $(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, $i = 1, 2$ são ambos desconhecidos. Proponha uma estatística de teste, baseada na forma quadrática

$$T^2 = \left(\frac{1}{n_1} + \frac{1}{n_2} \right)^{-1} (\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2 - \boldsymbol{\Delta})' \hat{\boldsymbol{\Sigma}}^{-1} (\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2 - \boldsymbol{\Delta}),$$

para testar as hipóteses $H_0 : \mathbf{R}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \boldsymbol{\Delta}$ vs $\mathbf{R}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \neq \boldsymbol{\Delta}$, em que $\mathbf{R}_{(c \times p)}$, $c \leq p$ e $\boldsymbol{\Delta}_{(c \times 1)}$ são matrizes conhecidas de posto linha completo. Encontre sua distribuição exata e mostre como calcular o valor crítico e o nível descritivo associado.

6. Utilize o teste F para a igualdade de vetores de médias, apresentado em classe, para testar as hipóteses $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ vs $H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$, no conjunto de dados da Iris considerando os grupos setosa e versicolor como grupos 1 e 2, respectivamente (com $\alpha = 5\%$). Considere que $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$. Utilize as variáveis *Petal.Length* e *Petal.Width*. Apresente os vetores de médias estimadas de cada grupo, as matrizes de covariâncias estimadas de cada grupo e a ponderada, o valor da estatística do teste, o nível descritivo e a respectiva conclusão. Caso você tenha rejeitado H_0 , identifique em qual(is) variável(is) reside a diferença, através da metodologia vista em sala de aula.
7. Utilize o teste de Box para a igualdade de matrizes de covariâncias, apresentado em classe, para testar as hipóteses $H_0 : \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3$ vs H_1 : pelo menos uma diferença, no conjunto de dados da Iris considerando os grupos setosa, versicolor e virginica como 1, 2 e 3, respectivamente. Utilize as variáveis *Petal.Length* e *Petal.Width*. Apresente as matrizes de covariâncias estimadas de cada grupo e a ponderada, o valor da estatística do teste, o nível descritivo e a respectiva conclusão. Apresente os vetores

de médias estimadas de cada grupo, o valor da estatística do teste, o nível descritivo e a respectiva conclusão.

8. Com relação ao modelo linear multivariado homocedástico, visto em sala, responda os itens:

a) Prove que o e.m.v de \mathbf{B} é dado por $\widehat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$.

b) Prove que o e.m.v de $\mathbf{\Sigma}$ é dado por $\widehat{\mathbf{\Sigma}} = \frac{1}{n} (\mathbf{Y} - \mathbf{X}\widehat{\mathbf{B}})' (\mathbf{Y} - \mathbf{X}\widehat{\mathbf{B}})$.

9. Considere o conjunto de dados USArrests dividido de acordo com a variável “regiao” (constante no programa “AnáliseCP ME 713 2S 2020.r”, disponível no site do curso). O objetivo é comparar os quatro grupos (regiões) com relação as quatro variáveis (Assassinato, Assalto, PMAU, VS), identificando todas as diferenças existentes. Analise esses dados, descritiva e inferencialmente, da forma mais completa possível, para alcançar esse objetivo. Descreva/mencione, de forma apropriada, todas as metodologias utilizadas, escreva de forma clara as conclusões e comentários, eventualmente ajustando modelos reduzidos (se pertinente). Apresente as análises em forma de relatório.
10. No item anterior, una as regiões “Nordeste” e “Sul” em “Leste”, bem com as regiões “Centro-Oeste” e “Oeste” em “Oeste”. Considere as quatro variáveis mencionadas. Compare os dois grupos, de forma apropriada, em relação as matrizes de covariâncias, bem com em relação aos vetores de médias. Verifique, de forma adequada, as suposições para a realização desses testes.
11. Considere os dados “measure”, disponíveis no pacote do R “MVA” (digite: `install.packages("MVA");library(MVA);demo("Ch-MVA")`), para acessar ao banco de dados. Nele se encontram medidas de peito, cintura e quadril em 20 indivíduos (em polegadas), de ambos os sexos. O objetivo é comparar os dois grupos, em relação as essas variáveis, identificando

todas as diferenças existentes. Analise esses dados, descritiva e inferencialmente, da forma mais completa possível, para alcançar esse objetivo. Descreva/mencione, de forma apropriada, todas as metodologias utilizadas, escreva de forma clara as conclusões e comentários, eventualmente ajustando modelos reduzidos (se pertinente). Apresente as análises em forma de relatório.

12. Nas Olimpíadas de 1988, em Seul, o heptatlo foi vencido por uma das estrelas do atletismo feminino dos EUA, Jackie Joyner-Kersey. Os resultados para todos os 25 competidores em todas as sete (inclusive o escore total), podem ser encontrados no arquivo “heptathlon” (digite: `install.packages("MVA");library(MVA);demo("Ch-MVA")`). As variáveis são: `hurdles` (resultados de 100 m com barreiras, em segundos), `highjump` (resultados de salto em altura, em metros), `shot` (resultados de arremesso de peso, em metros), `run200m` (resultados de 200 m rasos, em segundos), `longjump` (resultados de salto em distância, em metros), `javelin` (resultados de lançamento de dardos, em metros), `run800m` (resultados de 800 m rasos, em segundos), `score` (pontuação total). Objetivo é modelar, simultaneamente, “`run200m`” e “`run800m`”, em função de todas as outras covariáveis (use as covariáveis padronizadas). Analise esses dados, descritiva e inferencialmente, da forma mais completa possível, para alcançar esse objetivo. Descreva/mencione, de forma apropriada, todas as metodologias utilizadas, escreva de forma clara as conclusões e comentários, eventualmente ajustando modelos reduzidos (se pertinente). Apresente as análises em forma de relatório.