# SME0822 - Análise Multivariada e Aprendizado Não Supervisionado

Francisco Miranda

November 8, 2022

## Contents



# 1 Atividade 3 - Francisco Miranda - 4402962

```
# bibliotecas utilizadas
library(tidyverse)
library(ggExtra)
library(ascii)
library(multcomp)
options(asciiType = "org")
library(jocre)
library('rockchalk')
library(MVTests)
```

## 1.1 Exercício 2

Carregamos o conjunto de dados que representa uma amostra aleatória de tamanho 42 observada, da qual temos o vetor de médias e a matrix de variância covariância.

```
n <- 42
mu <- c(0.564, 0.603)
S <- matrix(c(0.0144, 0.0117, 0.0117, 0.0146), ncol = 2)
```

## 1.2 (b) região de confiança para o vetor de médias

Encontramos a região de 95% de confiança para $\mu$ com auxílio da função pacote ‘mvrnorm‘

```
data <- mvrnorm(n= n,
       mu=mu,
       Sigma=S
       )
```

```
cset(data, method = "hotelling", alpha = 0.05)
```

```
Parameter estimates and projected boundaries of the 2-dimensional
95% simultaneous confidence region:

     Estimate Lower Upper
[1,]    0.557 0.508 0.606
[2,]    0.595 0.548 0.642
```

## 1.3 (e) Testes de hipótese

```
test <- OneSampleHT2(data, mu=c(0.6,0.58))
summary(test)
```

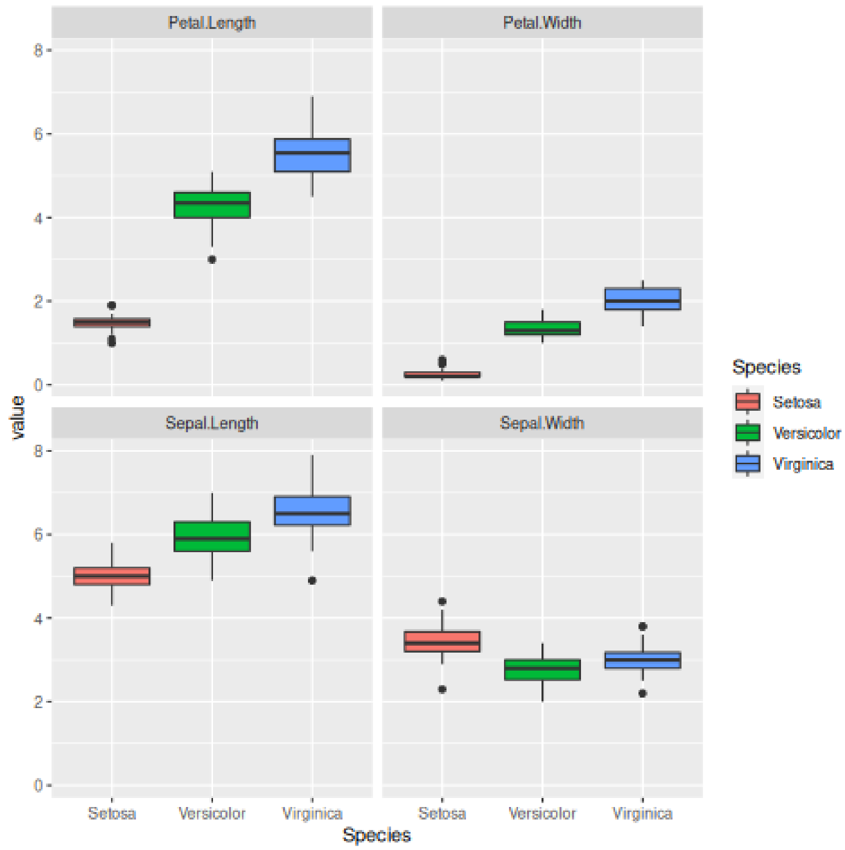
```
       One Sample Hotelling T Square Test

Hotelling T Sqaure Statistic = 27.03965
```

```
 F value = 13.19 , df1 = 2 , df2 = 40 , p-value: 3.99e-05

                 Descriptive Statistics

            [,1]       [,2]
N     42.0000000 42.0000000
Means  0.5572382  0.5952397
Sd     0.1238484  0.1181872


              Detection important variable(s)

      Lower     Upper  Mu0 Important Variables?
1 0.5080501 0.6064263 0.60                FALSE
2 0.5483000 0.6421794 0.58                FALSE
```

### 1.4 Exercício 4

Vamos avaliar se há diferença estatisticamente significante entre as variáveis do conjunto de dados ‘iris‘. Iniciamos nossa análise observando a distribuições dos tamanhos de sépala e pétala de acordo com a espécie.

```
iris |> pivot_longer(cols = !"Species") |>
  ggplot(aes(x = Species, y = value, fill = Species)) +
  geom_boxplot() +
  facet_wrap(~name)
```




Vamos tentar captar as diferenças significativas entre os três grupos através do teste MANOVA. Nossas hipóteses são:

$h_0$ não existem diferenças significativas entre as médias dos grupos, versus

$h_a$:* a média de pelo menos um dos grupos é diferente das demais.

```
fit <- manova(
  cbind(Sepal.Length, Sepal.Width, Petal.Length, Petal.Width)
     ~ Species, data = iris)

print(summary(fit))

           Df Pillai approx F num Df den Df    Pr(>F)
Species     2 1.1919   53.466      8    290 < 2.2e-16 ***
Residuals 147
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

O valor-p obtido rejeita a hipótese nula para um nível de significância de 5%. Logo, existem diferenças entre as médias das variáveis estudadas entre as espécies de flor.

### 1.5 Exercicio 5.9

Carregamos o conjunto de dados com medidas de 61 ursos do Alaska.

```
M <- matrix(c(3266.46, 1343.97, 731.54, 1175.50, 162.68, 238.37,
              1343.97, 721.91, 324.25, 537.35, 80.17, 117.73,
              731.54, 324.25, 179.28, 281.17, 39.15, 56.80,
              1175.50, 537.35, 281.17, 474.98, 63.73, 94.85,
              162.68, 80.17, 39.15, 63.73, 9.95, 13.88,
              238.37, 117.73, 56.80, 94.85, 13.88, 21.26),
            nrow = 6, ncol = 6, byrow = T)
colnames(M) <- c("Peso","Comprimento Corporal","Pescoco",
                 "Perímetro", "Comprimento Cabeca", "Largura Cabeca")
mu <- c(95.52, 164.38, 55.69, 93.39, 17.98, 31.13)
n <- 61
p <- 6
```

#### 1.5.1 (a) Intervalos de 95% de confiança simultâneos para as seis medidas

```
# Intervalos  simulatânea de 95% de confiança

crit <- qf(0.05, p, n - p, lower.tail = F)
# Máximo erro
E <- sqrt((n-1)*p*crit/(n-p))* sqrt(diag(M)/n)
# intervals
LI = mu -  E
LS = mu +  E

Res=cbind(mu, LI, LS)
colnames(Res)=c("D_bar", "LI","LS")
print(Res, digits = 3, type = "org")
```

```
     D_bar    LI    LS
[1,]  95.5  67.3 123.7
[2,] 164.4 151.1 177.6
[3,]  55.7  49.1  62.3
[4,]  93.4  82.6 104.1
[5,]  18.0  16.4  19.5
[6,]  31.1  28.9  33.4
```

### 1.5.2 (c) Intervalos de 95% de confiança de Bonferroni simultâneos para as seis medidas

```
##################################################
# Intervalos  de 95% de confiança de Bonferroni
##################################################

crit <- qt(0.05/(2*4), n-1, lower.tail = F)
# Máximo erro
E <- crit* sqrt(diag(M)/n)
# intervals
LI = mu -  E
LS = mu +  E

Res=cbind(mu, LI, LS)
colnames(Res)=c("D_bar", "LI","LS")
print(Res,digits=3, type = "org")
```

```
     D_bar    LI    LS
[1,]  95.5  76.7 114.4
[2,] 164.4 155.5 173.2
[3,]  55.7  51.3  60.1
[4,]  93.4  86.2 100.6
[5,]  18.0  16.9  19.0
[6,]  31.1  29.6  32.7
```