


1. Foi lido e introduzido no  o conjunto de dados “Remocao-fruto.txt”. Referem-se ao resultado da remoção de ervas daninhas no plantio de tomate. Uma área foi subdividida em parcelas de tamanhos iguais. Numas parcelas (sem remoção) foi permitido que as plantas daninhas crescessem livremente associadas aos tomateiros, enquanto nas outras (remoção) as ervas eram sistematicamente removidas, de modo que não se estabelecessem. Após um determinado tempo, um único fruto foi escolhido aleatoriamente de um pé e pesado (gramas). Em seguida, a planta foi removida para obtenção do peso seco das raízes (em gramas). Utilize o *output* que se encontra no Anexo para responder, sempre que possível às seguintes questões:
 - a) Classifique as variáveis que estão naquele **data frame**.
 - b) Complete os valores, **A**, **B**, **C** e **D** em falta no *output*.
 - c) Faça uma análise descritiva do peso dos frutos colhidos, separando as situações "remocao" e "naoremocao".
 - d) Analise o gráfico com os *boxplots* para descrever o peso dos frutos em função dos tratamentos e outro para o peso da raiz na mesma situação. Identifique cada um deles.
 - e) Calcule o intervalos de confiança a 95% para o peso dos frutos do tratamento sem remoção e interprete-o.
 - f) Teste a hipótese de que o peso médio dos frutos no tratamento com remoção é maior que no tratamento sem remoção. Responda de forma completa à pergunta.
 - g) Indique uma estimativa da proporção de parcelas em que foi efectuada remoção e o peso do tomate é superior a 60 gramas.
2. Sabe-se que a morte dos indivíduos da espécie *Buteo buteo* (águia de asa redonda) encontrados em campos de forragem e de beterraba sacarina pode estar relacionada com o uso de carbofuran. Numa determinada área geográfica, o estudo da concentração de carbofuran (mg/kg) revelou que 5% dos indivíduos da população apresentavam valores daquela concentração superiores a 3 mg/kg. Admita que a concentração de carbofuran tem distribuição normal de valor médio 1.2 mg/kg.
 - a) Determine a variância da variável em estudo.
 - b) Determine os limites do intervalo de concentrações centrado no valor médio que contém 50% dos casos.
 - c) Selecionando ao acaso uma amostra de 20 aves, qual a probabilidade de:
 - i) A concentração média amostral ultrapassar o valor 2.2 mg/kg?
 - ii) Encontrar mais de 4 aves com concentração de carbofuran superior a 3 mg/kg?

3. O modelo gama tem grandes aplicações em tempos de vida de organismos. Seja X uma variável aleatória seguindo um modelo gama, dependendo de dois parâmetros desconhecidos $\alpha > 0$ e $\beta > 0$, com função densidade definida como:

$$f(x|\alpha, \beta) = \frac{x^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha} \exp\left(-\frac{x}{\beta}\right), \quad \text{se } x > 0, \quad \text{e nula nos restantes valores de } x$$


Considere que se tem uma amostra aleatória de dimensão (X_1, X_2, \dots, X_n) associada a X .

- Sabendo que $E[X] = \alpha\beta$ e $Var[X] = \alpha\beta^2$, determine estimadores de α e β pelo método dos momentos.
- Nesta alínea considere $\alpha = 2$. Calcule o estimador de máxima verosimilhança para β .
Nota: Recorde que $\Gamma(n) = (n-1)!$, se $n \in \mathbb{N}$
- No seguinte *output*, `xval` identifica os valores de uma amostra de dimensão 20. Calcule as estimativas de α usando os dois estimadores obtidos na alínea anterior.

```
> xval

      [1]  4.246  4.901  1.799  7.095  6.990  1.461  0.609  6.880 11.086  7.658
     [11] 13.430 12.404  3.292  4.259  9.831  1.990  2.750  6.812  9.456  6.988

sum(xval)
[1] 123.937
> sum(xval*xval)
[1] 1031.891
```

- Um campo experimental foi utilizado para testar o crescimento de duas espécies florestais A e B . Analisaram-se 200 árvores da espécie A com 2 anos de idade, tendo-se obtido uma altura média de 145 cm e um desvio padrão de 15 cm. Uma amostra de 150 árvores da espécie B , com a mesma idade, conduziu a uma altura média de 141 cm e a um desvio padrão de 12 cm.
 - Indique uma estimativa para a diferença das alturas das duas espécies ao fim de 2 anos, assim como para a variância da altura de cada uma das espécies.
 - Será que em média as árvores da espécie A são mais altas do que as árvores da espécie B ao fim de 2 anos?
- Considere as variáveis aleatórias $X \sim \text{Exp}(1)$ e $S_k = \sum_{i=1}^k X_i$, $k = 2, 5, 10$ com $\{X_i\}$ independentes com a distribuição de X .
 - Escreva um comando em  para obter, por simulação, uma amostra de dimensão $n = 100$ com a distribuição de X .
 - Escreva os comandos necessários para obter, por simulação, uma amostra de dimensão $n = 100$ de cada uma das v.a. S_2 , S_5 e S_{10} .
- Diga, justificando convenientemente, se são **verdadeiras** ou **falsas** as afirmações em cada uma das seguintes alíneas. Corrija as que considerar falsas.
 - Seja χ_α , com $\alpha \in (0, 1)$, o quantil de probabilidade α de uma distribuição $\mathcal{N}(0, 1)$.
Tem-se $\chi_\alpha = -\chi_{1-\alpha}$
 - Se $X \sim \mathcal{N}(\mu, \sigma)$, então dada a amostra aleatória (X_1, \dots, X_n) retirada de X tem-se $\bar{X} \sim \mathcal{N}(\mu, \sigma)$.

ANEXO

Pergunta 1

```
> tomate<-read.table("Remocao-fruto.txt",header=TRUE,dec=".",sep=";",as.is=TRUE,na.strings=NA)
> attach(tomate)

> head(tomate)
      Raiz Fruto Tratamento
1 6.225 59.77 NaoRemocao
2 6.487 60.98 NaoRemocao
3 4.919 14.73 NaoRemocao
4 5.130 19.28 NaoRemocao
5 5.417 34.25 NaoRemocao
6 5.359 35.53 NaoRemocao

> dim(tomate)
[1] 40  A

> summary(Fruto)
      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      7.05   36.98   54.23   53.04   71.57   87.73

> by(Fruto,Tratamento,summary)
Tratamento: NaoRemocao
      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      14.73   33.77   54.23   50.88   64.46   87.73
-----
Tratamento: Remocao
      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      7.05   39.85   56.52   55.19   74.91   84.37
>
> by(Raiz,Tratamento,summary)
Tratamento: NaoRemocao
      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      4.426   5.402   6.119   6.053   6.598   7.614
-----
Tratamento: Remocao
      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      5.253   7.039   7.958   7.709   8.558   9.039

> (fruto.no.rem<-Fruto[Tratamento=="NaoRemocao"])
[1] 59.77 60.98 14.73 19.28 34.25 35.53 87.73 63.21 24.25 64.34 52.92 32.35
[13] 53.61 54.86 64.81 73.24 80.64 18.89 75.49 46.73

> (fruto.com.rem<-Fruto[Tratamento=="Remocao"])
[1] 7.05 38.94 60.77 84.37 70.11 14.95 70.70 80.31 82.35 20.07 73.79 50.08
[13] 78.28 41.48 37.47 40.15 52.26 46.64 71.01 83.03
>
> sort(fruto.no.rem)
[1] 14.73 18.89 19.28 24.25 32.35 34.25 35.53 46.73 52.92 53.61 54.86 59.77
[13] 60.98 63.21 64.34 64.81 73.24 75.49 80.64 87.73
```

```
> shapiro.test(fruto.no.rem)
```

Shapiro-Wilk normality test

data: fruto.no.rem

W = 0.9537, p-value = 0.4271

```
> shapiro.test(fruto.com.rem)
```

Shapiro-Wilk normality test

data: fruto.com.rem

W = 0.9213, p-value = 0.1048

```
> shapiro.test(fruto.no.rem-fruto.com.rem)
```

Shapiro-Wilk normality test

data: fruto.no.rem - fruto.com.rem

W = 0.9534, p-value = 0.4217

```
> var.test(fruto.no.rem,fruto.com.rem)
```

F test to compare two variances

data: fruto.no.rem and fruto.com.rem

F = 0.8261, num df = 19, denom df = 19, p-value = 0.6814

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.3269861 2.0871370

sample estimates:

ratio of variances

0.8261142

```
> t.test(fruto.no.rem)
```

One Sample t-test

data: fruto.no.rem

t = 10.4575, df = 19, p-value = 2.548e-09

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval:

40.69699 B

sample estimates:

mean of x

50.8805

```
> t.test(fruto.no.rem,fruto.com.rem,var.equal=TRUE,alternative="less")
```

Two Sample t-test

data: fruto.no.rem and fruto.com.rem

t = -0.5958, df = 38, p-value = 0.2774

alternative hypothesis: true difference in means is less than 0

95 percent confidence interval:

```

-Inf 7.885881
sample estimates:
mean of x mean of y
  50.8805   55.1905

> t.test(fruto.no.rem,fruto.com.rem,alternative="less",paired=TRUE)

```

Paired t-test

```

data:  fruto.no.rem and fruto.com.rem
t = -0.552, df = 19, p-value = 0.2937
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
 -Inf 9.192218
sample estimates:
mean of the differences
      C

```

```

> t.test(fruto.no.rem,fruto.com.rem,paired=TRUE)

```

Paired t-test

```

data:  fruto.no.rem and fruto.com.rem
t = -0.552, df = 19, p-value = 0.5874
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -20.65372  12.03372

```

