Instituto Superior de Agronomia
Modelos Matemáticos – 2015/2016

Teste do Módulo I-2ª data 29 de Abril de 2016

1. Após a colheita, as peras colhidas no pomar são separadas de acordo com o seu calibre (em mm). Da colheita de um dia retiraram-se 1000 peras cujos calibres se encontram classificados na seguinte tabela:

Classes de calibre	[30;50]	]50;55]	]55;60]	]60;70]	]70;90]
Nº de peras	40	160	390	350	60

Nota: Utilize os resultados que lhe sejam úteis, apresentados no Anexo, para responder às seguintes questões.

a) Identifique e classifique, justificando, a variável em estudo.

b) Elabore uma tabela de frequências relativas e esboce a sua representação gráfica.

c) Determine a média e a mediana, aproximadas, do calibre das peras. Comente.

d) Indique uma estimativa da proporção de peras com calibre superior a 60 mm. Qual a precisão associada a esta estimativa com uma confiança de 95%?

e) Poderá o produtor afirmar que, em média, o calibre das peras deste pomar é superior a 60 mm? Justifique convenientemente.

2. O modelo de Rayleigh tem sido usado para caracterizar situações ambientais em que poderão ocorrer valores "severos", como por exemplo altura de ondas e rajadas de vento. Vamos considerar X uma variável aleatória seguindo um modelo de Rayleigh numa forma muito simples, apenas com um parâmetro desconhecido $\theta > 0$, com função densidade definida como:

$$f(x|\theta) = \frac{x}{\theta} \exp[-x^2/2\theta], \quad \text{se } \ x > 0, \quad \text{e nula nos restantes valores de } x$$

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

a) Calcule o estimador de θ pelo método dos momentos.
Nota: Sabe-se que $E[X] = \sqrt{\frac{\theta\pi}{2}}$ e $Var[X] = \frac{4-\pi}{2}\theta$.

b) Calcule o estimador de máxima verosimilhança para θ.

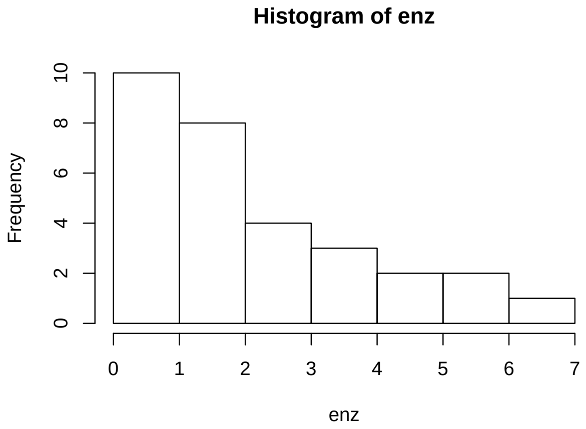
3. Num delineamento completamente casualizado foram estudados dois métodos diferentes de cultivo de milho tendo sido regitados os rendimentos por ha em diversas parcelas de terreno onde os dois métodos foram aplicados. Os resultados obtidos, bem como algumas análises encontram-se no Anexo.

a) Indique estimativas para o valor médio e variância do rendimento por ha em cada método?

b) Os dados recolhidos permitem afirmar que algum dos métodos é o melhor? Justifique de forma completa.

4. Suponha que está a medir os tempos de sobrevivência de uma enzima numa dada solução e se obtiveram os valores (em horas) apresentados no Anexo.

Pretende-se modelar a distribuição de probabilidade do tempo de vida destas enzimas. Numa primeira fase fez-se o histograma das observações, que se apresenta a seguir



a) Eu dir-lhe-ia que suspeito que seja um modelo exponencial de parâmetro $\beta > 0$, i.e.
$f(x) = \frac{1}{\beta}\exp(-x/\beta)$ para $x > 0$ e $F(x) = 1 - \exp(-x/\beta)$ para $x > 0$.
Comente esta minha suspeição.

b) Vamos agora ver se eu terei razão!

   i) Com base nos dados da amostra indique uma estimativa do parâmetro β.

   ii) Considere os dados agrupados nas seguintes classes
   $[0, 1]$ $]1, 2]$ $]2, 3]$ $]3, 5]$ $]5, \infty[$

   Teste agora a hipóteses de os tempos destas enzimas seguirem uma distribuição exponencial. Justifique convenientemente a sua resposta.
   Nota: Utilize os resultados que considere úteis, apresentados no Anexo.

5. Sejam $(X_1, X_2, \ldots, X_n)$ e $(Y_1, Y_2, \ldots, Y_n)$ duas amostras aleatórias de dimensão $n \geq 2$ retiradas de populações normais independentes, $X \frown \mathcal{N}(\mu_1, \sigma)$ e $Y \frown \mathcal{N}(\mu_2, \sigma)$. Suponha que σ é conhecido.

Pretende-se testar a hipótese $H_0 : \mu_1 = \mu_2$ *vs* $H_0 : \mu_1 \neq \mu_2$ e decide-se usar a seguinte regra de decisão:

- Se os intervalos de confiança com nível α $(0 < \alpha < 1)$ tiverem pelo menos um valor em comum, i.e., estiverem sobrepostos, a decisão é *não rejeitar H_0*;
- Se os intervalos não se sobrepuserem *rejeita-se H_0 a favor de H_1*.

a) Se H_0 verdadeira mostre que a probabilidade de que os dois intervalos **se sobreponham** é $2\Phi(\sqrt{2}\ z_{\alpha/2}) - 1$.
Nota: Tenha em conta que os dois intervalos de confiança $(1-\alpha) \times 100\%$ para os valores médios de X e de Y não se sobrepõem quando a diferença entre as médias amostrais for superior à soma das semi-amplitudes de cada um dos intervalos.

b) Sendo $\alpha = 0.05$ indique o valor da probabilidade calculada na alínea anterior.

c) Usando a regra de decisão indicada qual deverá ser o valor de α para que a probabilidade do erro tipo I seja 0.05?

ANEXO

```
###----------------
####exercício 1
###---------------


> hist(amostra,breaks=c(30,50,55,60,70,90),plot=F)
$breaks
[1] 30 50 55 60 70 90

$counts
[1]  40 160 390 350  60

$density
[1] 0.002 0.032 0.078 0.035 0.003

$mids
[1] 40.0 52.5 57.5 65.0 80.0

$equidist
[1] FALSE


> t.test(amostra)

        One Sample t-test

data:  amostra
t = 232.2, df = 999, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 59.87518 60.89582
**********************

> t.test(amostra, alternative="greater",mu=60)

        One Sample t-test

data:  amostra
t = 1.4824, df = 999, p-value = 0.06928
alternative hypothesis: true mean is greater than 60
95 percent confidence interval:
 59.95734      Inf
*****************

> t.test(amostra,alternative="greater")

        One Sample t-test

data:  amostra
t = 232.2, df = 999, p-value < 2.2e-16
alternative hypothesis: true mean is greater than 0
95 percent confidence interval:
 59.95734      Inf
**************
```

```
###------------
#### Exercício 3
#####-----------


> rend1
 [1] 83 91 94 89 89 96 91 92 90 91 90 81 83 84 83 88 91
> rend2
 [1]  89  84 101 100  91  93  96  95  94  78  82  81  77  79  81  80  81
> shapiro.test(rend1)

        Shapiro-Wilk normality test

data:  rend1
W = 0.9203, p-value = 0.1492

> shapiro.test(rend2)

        Shapiro-Wilk normality test

data:  rend2
W = 0.898, p-value = 0.06283

> shapiro.test(rend1-rend2)

        Shapiro-Wilk normality test

data:  rend1 - rend2
W = 0.9711, p-value = 0.8368

> var.test(rend1,rend2)

        F test to compare two variances

data:  rend1 and rend2
F = 0.2819, num df = 16, denom df = 16, p-value = 0.01561
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1020840 0.7784008
sample estimates:
ratio of variances
         0.2818905

> t.test(rend1,rend2)

        Welch Two Sample t-test

data:  rend1 and rend2
t = 0.6345, df = 24.356, p-value = 0.5317
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.176887  6.000416
sample estimates:
mean of x mean of y
 88.58824  87.17647
```

```
> t.test(rend1,rend2,paired=TRUE)

        Paired t-test

data:  rend1 and rend2
t = 0.8524, df = 16, p-value = 0.4066
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.099274  4.922804
sample estimates:
mean of the differences
               1.411765

> wilcox.test(rend1,rend2)

        Wilcoxon rank sum test with continuity correction

data:  rend1 and rend2
W = 166, p-value = 0.4681
alternative hypothesis: true location shift is not equal to 0

Warning message:
In wilcox.test.default(rend1, rend2) :
  cannot compute exact p-value with ties
```

```
##---------
## Exercício 4
##---------


> enz<-c(4.75, 3.4, 1.8, 0.9, 2.2, 0.4, 5.8, 1.6, 2.4, 2.25,
+ 1.36, 0.93, 3.97, 6.95, 1.24, 0.80, 0.30, 0.02, 1.04, 0.89,
+ 3.82,1.25,0.04,2.65,4.3,0.52,5.8, 0.12, 1.5, 1.01)

> hist(enz)

> sum(enz)
[1] 64.01

> beta.est<-mean(enz)
> hist(enz,breaks=c(0,1,2,3,5,7),plot=F)
$breaks
[1] 0 1 2 3 5 7

$counts
[1] 10  8  4  5  3



$mids
[1] 0.5 1.5 2.5 4.0 6.0

$xname
[1] "enz"

$equidist
[1] FALSE


> ni<-hist(enz,breaks=c(0,1,2,3,5,7),plot=F)$counts



> p1<-pexp(1,1/beta.est)
> p2<-pexp(2,1/beta.est)-pexp(1,1/beta.est)
> p3<-pexp(3,1/beta.est)-pexp(2,1/beta.est)
> p4<-pexp(5,1/beta.est)-pexp(3,1/beta.est)
> p5<-1-pexp(5,1/beta.est)
> prob<-c(p1,p2,p3,p4,p5)
> prob
[1] 0.37417016 0.23416685 0.14654860 0.14911215 0.09600223
> prob*length(enz)
[1] 11.225105  7.025006  4.396458  4.473365  2.880067
>
```