Instituto Superior de Agronomia
Modelos Matemáticos e Aplicações – 2016/2017

Teste do Módulo I– duração: 2h30 15 de Março de 2017

1. A espécie de borboletas *Heliothis Armigera* é considerada uma praga que ataca diversas culturas. Os ovos são arredondados, amarelo-acastanhados e aparecem dispostos em grupo. Numa cultura atacada por este tipo de insecto foram recolhidas 7640 folhas e contou-se o número de ovos em cada folha.

Os dados observados foram introduzidos no R, tendo-se obtido o *output* que se encontra no Anexo e que deve utilizar para responder às seguintes questões:

a) Complete os valores referidos por A, B, C, D e E, em falta no *output*.

b) Compare, comentando, os valores da moda, mediana e média do número de ovos obervados por folha.

c) Faça o esboço de um *boxplot* para os dados observados.

d) Indique uma estimativa da proporção de folhas onde foram encontrados ovos.

e) Indique um intervalo de confiança a 95% para a proporção de folhas com ovos. Comente.

f) Poder-se-á considerar que a percentagem de folhas com ovos é superior a 10% ? Justifique convenientemente.

g) Indique um intervalo de confiança a 95% para o número médio de ovos por folha.

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 velocidade de vento. Vamos considerar X uma variável aleatória seguindo um modelo de Rayleigh apenas com um parâmetro desconhecido $\theta > 0$, com função densidade definida como:

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

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

Nota: Sabe-se que $E[X] = \theta\sqrt{\frac{\pi}{2}}$ e $Var[X] = \frac{4-\pi}{2}\theta^2$.

a) Obtenha o estimador de θ pelo método dos momentos.

b) Verifique se o estimador dos momentos obtido na alínea a) é um estimador centrado de θ.

c) Obtenha o estimador de máxima verosimilhança para θ.

d) Considere os valores observados da seguinte amostra, extraída daquela população, com a qual se realizaram os cálculos apresentados:

```
> amostra
 [1] 3.47 4.18 2.08 3.03 3.39 2.89 1.84 0.56 3.04 2.28 2.31 1.96 3.29 3.60 4.56
[16] 2.49 4.61 2.18 1.53 2.28 1.98 1.24 2.24 4.58 2.74 2.61 1.65 4.58 3.74 3.21
> sum(amostra)
[1] 84.14
> sum(amostra*amostra)
[1] 268.2656
```

i) Determine estimativas para θ.

ii) Determine uma estimativa do coeficiente de variação de X, CV_X.

3. Durante 15 dias fizeram-se medições diárias do pH da água à entrada e à saída de um tanque natural de tratamento de águas residuais. Os dados foram introduzidos no *software* R. Utilize, sempre que possível, os resultados apresentados no Anexo para responder às seguintes questões (alguns cálculos apresentados são desnecessários ou inadequados).

a) Determine uma estimativa para a diferença média entre os valores do pH da água à entrada e à saída do tanque.

b) Os dados recolhidos permitem concluir que a passagem da água pelo tanque conduz à alteração média diária do valor do pH? Responda de forma completa à questão.

4. Considere duas variáveis aleatórias contínuas X e Y, independentes, com distribuição uniforme em $[0, 1]$. Escreva os comandos que lhe permitem:

a) gerar dois vectores, x e y extraídos de X e de Y, respectivamente, com 1000 valores cada;

b) obter o vector $z = x + y$;

c) desenhar o histograma de z;

d) Interprete os seguintes comandos. Que conclui?

```
> ks.test(z,"punif",0,2)
        One-sample Kolmogorov-Smirnov test
data:  z
D = 0.1343, p-value = 4.441e-16
alternative hypothesis: two-sided


> ks.test(z,"pnorm",mean(z),sd(z))
        One-sample Kolmogorov-Smirnov test
data:  z
D = 0.0321, p-value = 0.2529
alternative hypothesis: two-sided
```

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 $H_0 : \mu_1 = \mu_2 \quad vs \quad H_0 : \mu_1 \neq \mu_2$ e decide-se usar a seguinte regra de decisão:

- Se os intervalos de confiança com nível $\alpha \quad (0 < \alpha < 1)$ 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) Sendo 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 I

```
########  Pergunta 1   ########

> ovos<-c(rep(0,6685),rep(1,715),rep(2,148),rep(3,55),
+ rep(4,18),rep(5,9),rep(6,6),rep(7,2),rep(8,1),rep(9,1))
>
> ni<-table(ovos); ni
ovos
   0    1    2    3    4    5    6    7    8    9
6685  715  148   55   18    9    6    2    1    1

> A <-cumsum(ni); fi<-round(ni/sum(ni),4); Fi<-round(Ni/sum(ni),4); xi<-seq(0,9)
>
> cbind(xi,ni,Ni,fi,Fi)
  xi   ni   Ni     fi     Fi
0  0 6685 6685 0.8750 0.8750
1  1  715 7400 0.0936   B
2  2  148 7548 0.0194 0.9880
3  3   55 7603 0.0072 0.9952
4  4   18 7621 0.0024 0.9975
5  5    9 7630 0.0012 0.9987
6  6    6 7636 0.0008 0.9995
7  7    2 7638 0.0003 0.9997
8  8    1 7639 0.0001 0.9999
9  9    1 7640 0.0001 1.0000


> library(fBasics); basicStats(ovos)
                   ovos
nobs        7640.000000
NAs            0.000000
Minimum        0.000000
Maximum        9.000000
1. Quartile    0.000000
3. Quartile    0.000000
Mean           0.178010
Median           C
Sum         1360.000000
SE Mean          D
LCL Mean       0.165099
UCL Mean         E
Variance       0.331445
Stdev          0.575712
Skewness       5.136906
Kurtosis      38.921557


> xtot<-7640-6685


> binom.test(x=xtot, n=7640, p = 0.1,alternative =  "greater")

        Exact binomial test
data:  xtot and 7640
number of successes = 955, number of trials = 7640, p-value = 1.056e-12
alternative hypothesis: true probability of success is greater than 0.1
95 percent confidence interval:
 0.1188175 1.0000000
sample estimates:
probability of success
                 0.125
```

```
> binom.test(x=xtot, n=7640, p = 0.1)

        Exact binomial test
data:  xtot and 7640
number of successes = 955, number of trials = 7640, p-value = 1.86e-12
alternative hypothesis: true probability of success is not equal to 0.1
95 percent confidence interval:
 0.1176628 0.1326246
sample estimates:
probability of success
                 0.125


> binom.test(x=xtot, n=7640, p = 0.1,alternative =  "less")

        Exact binomial test
data:  xtot and 7640
number of successes = 955, number of trials = 7640, p-value = 1
alternative hypothesis: true probability of success is less than 0.1
95 percent confidence interval:
 0.0000000 0.1313951
sample estimates:
probability of success
                 0.125


########  Pergunta 3   ########

> dia <- seq (1:15)
> entrada <- c(7.0,7.6,7.3,7.0,7.1,7.5,7.1,7.0,7.0,7.5,7.4,7.4,6.8,7.0,7.9)
> saida <- c(7.5,7.9,7.3,7.6,7.4,8.0,7.4,7.7,7.9,7.6,7.9,7.3,7.7,7.1,7.5)

> shapiro.test(entrada)
          Shapiro-Wilk normality test
       data:  entrada
       W = 0.9168, p-value = 0.1722

> shapiro.test(saida)
        Shapiro-Wilk normality test
    data:  saida
    W = 0.9587, p-value = 0.6695

> shapiro.test(entrada-saida)
               Shapiro-Wilk normality test
       data:  entrada - saida
       W = 0.9695, p-value = 0.8509

> t.test(entrada,saida,paired=FALSE,var.equal=TRUE)
               Two Sample t-test
       data:  entrada and saida
       t = -3.3466, df = 28, p-value = 0.002343
       alternative hypothesis: true difference in means is not equal to 0
       95 percent confidence interval:
        -0.5588540 -0.1344794
       sample estimates:
       mean of x mean of y
        7.240000  7.586667

> t.test(entrada,saida,paired=TRUE)
               Paired t-test
       data:  entrada and saida
       t = -3.6665, df = 14, p-value = 0.002540
       alternative hypothesis: true difference in means is not equal to 0
       95 percent confidence interval:
        -0.5494560 -0.1438773
       sample estimates:
       mean of the differences
                   -0.3466667
```