Instituto Superior de Agronomia
Modelos Matemáticos – 2015/2016

Teste do Módulo I 6 de Abril de 2016

1. O *Scaphoideus titanus* é um insecto que desenvolve todo o seu ciclo de vida na videira e tem uma geração por ano. As larvas abrem os seus ovos, que foram depositados na casca das cepas de Agosto a Outubro do ano anterior, a partir de Maio surgindo o insecto que ao picar as videiras as vai infectando. Um dos primeiros sinais de doença da videira é as folhas começarem a enrolar sobre a página inferior e a sobrepor-se. Foram observadas 525 videiras e contou-se, em cada uma, o número de folhas com sinal de doença.

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

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

b) Determine as seguintes características amostrais: moda, mediana e média do número de folhas doentes por videira. Compare os valores obtidos. Alguns valores estão representados por letras no Anexo, que já terá calculado em a).

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

d) Indique uma estimativa da proporção de videiras com folhas doentes.

e) Calcule um intervalo de confiança a 95% para a proporção de videiras com folhas doentes. Comente.

f) Indique um intervalo de confiança a 95% para o número médio de folhas doentes por videira.

g) Poder-se-á dizer que, em média, há mais de 1 folha doente por videira? Justifique.

2. Supõe-se que a concentração, X, de dióxido de enxofre (SO_2) (em partes por milhão) numa certa cidade é modelada por uma lei cuja função densidade é assim definida:

$$f(x|\theta) = \frac{x^{-1/2}}{\sqrt{2\pi\theta^2}}\, exp\left(-\frac{x}{2\theta^2}\right), \quad \text{se } \; x > 0, \quad \theta > 0, \quad \text{e } \; \text{nula nos restantes valores de } x$$

Tem-se $E[X] = \theta^2$ e $E[X^2] = 3\theta^4$.

Seja $(X_1, X_2, .., X_n)$ uma amostra aleatória de dimensão n associada a X.

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

b) Calcule o estimador de máxima verosimilhança para θ. Será o estimador de máxima verossimilhança de θ centrado?

c) No seguinte *output*, `xval` identifica os valores de uma amostra de dimensão 25. Calcule as estimativas de θ usando os dois estimadores obtidos nas alíneas anteriores.

```
> xval
 [1]  0.043  3.529  0.038  1.943  0.837  0.034 22.706  1.049  3.882
[10]  1.089  6.096  8.966  0.803  6.400  0.055  0.383 15.953  1.074
[19]  7.120  0.283  0.039  0.004  1.353  8.733  0.138

> sum(xval)
[1] 92.55
> sum(xval*xval)
[1] 1093.7
```

d) Indique uma estimativa para o coeficiente de variação, $CV = \frac{\sigma}{\mu} \times 100\%$.

3. Seja X uma v.a. que representa as alturas, em metros, de uma espécie de arbustos, que se admite ter distribuição normal, $X \frown \mathcal{N}(\mu, \sigma)$.

a) Nesta alínea considere $Z = \dfrac{X - \mu}{\sigma}$, portanto a variável Z é normal, $Z \frown \mathcal{N}(0, 1)$. Quais os comandos em R necessários para obter os gráficos da função densidade de Z e de Z^2?

b) Considere agora $\mu = 0$ m e $\sigma^2 = 0.4$ m^2. Estamos interessados em estudar uma característica daqueles arbustos que é modelada por uma expressão em que aparece X^2.
Escreva os comandos necessários para calcular, com recurso ao R, $P[a < X^2 < b]$.

4. Na tabela de contingência que a seguir se apresenta, 353 amostras de água do mar são classificadas segundo dois factores:

<u>Factor 1</u>: distância da costa à qual foram recolhidas as amostras de água (A – menor que 5 km; B – entre 5 e 15 km e C – mais do que 15 km)

<u>Factor 2</u>: nível de mercúrio detectado (I – irrelevante; II – detectável, mas abaixo dos níveis perigosos e III – acima dos níveis perigosos).

	Nível de mercúrio		
Distância	I	II	III
A	23	47	53
B	29	44	41
C	32	45	39

a) Considere apenas as distâncias a mais de 15km da costa, ou seja, <u>o nível C do factor distância</u>. Apenas para estas distâncias será razoável admitir que os níveis de mercúrio apresentam a mesma distribuição pelas 3 categorias? Justifique.

b) Face a estes dados, considera que a proximidade da costa influencia os níveis de mercúrio na água? (considere $\alpha = 0.05$).
Nota: para responder a esta alínea considere os resultados apresentados no Anexo. Calcule o valor de F, em falta.

ANEXO

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

> folhas<-c(rep(0,235),rep(1,161),rep(2,58),rep(3,35),
+ rep(4,18),rep(5,9),rep(6,6),rep(7,2),rep(8,1))
>
> ni<-table(folhas); ni
folhas
  0   1   2   3   4   5   6   7   8
235 161  58  35  18   9   6   2   1
>
> Ni<-cumsum(ni); fi<-round(ni/sum(ni),3); Fi<-round(Ni/sum(ni),3)
> xi<-seq(0,8)
>
> cbind(xi,ni,Ni,fi,Fi)
  xi  ni  Ni    fi    Fi
0  0 235 235 0.448 0.448
1  1 161 396 0.307 0.754
2  2  58 454   A   0.865
3  3  35 489 0.067 0.931
4  4  18 507 0.034 0.966
5  5   9 516 0.017 0.983
6  6   6 522 0.011   B
7  7   2 524 0.004 0.998
8  8   1 525 0.002 1.000
>
> mean(folhas)
[1] 1.060952
>
> library(fBasics)
> basicStats(folhas)
                folhas
nobs        525.000000
NAs           0.000000
Minimum       0.000000
Maximum       8.000000
1. Quartile   0.000000
3. Quartile   1.000000
Mean          1.060952
Median          C
Sum         557.000000
SE Mean       0.060533
LCL Mean        D
UCL Mean      1.179870
Variance      1.923759
Stdev           E
Skewness      1.788621
Kurtosis      3.549154
```

```
#### Exercício 4

> respostas<-matrix(c(23,47,53,29,44,41,32,45,39),byrow=T,nr=3,
+ dimnames = list(c("A", "B","C"),c("I", "II", "III")))
> respostas
   I II III
A 23 47  53
B 29 44  41
C 32 45  39

> chisq.test(respostas)

        Pearson's Chi-squared test

data:  respostas
X-squared = 3.7294, df = 4, p-value = F

> chisq.test(respostas)$expected
         I       II      III
A 29.26912 47.38810 46.34278
B 27.12748 43.92068 42.95184
C 27.60340 44.69122 43.70538

> (chisq.test(respostas)$residuals)^2
          I           II        III
A 1.3427765 0.0031785014 0.95632226
B 0.1292540 0.0001432510 0.08869666
C 0.7002796 0.0021334447 0.50658804
```