Instituto Superior de Agronomia
Modelos Matemáticos – 2016/2017

Tópicos de uma resolução do Teste do Módulo I 15 de Março de 2017

1. a) Vamos então indicar os valores em falta no *output* (vale a pena introduzir no R e tentar executar os comandos, para ver e compreender os resultados)

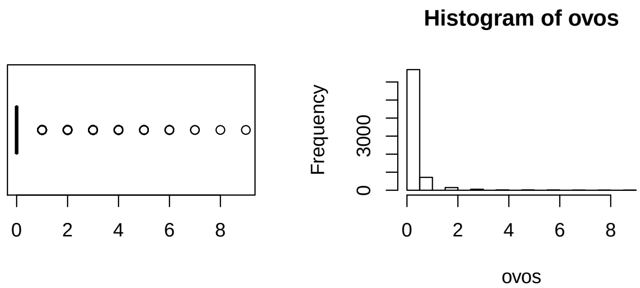
```
A  Ni
B  0.9686
C  0  a mediana é >= 1o quartil e <= 3o quartil, portanto só pode ser =0
D  é dado por s/sqrt(n)=0.006587
E  UCL Mean= média -(semi-amplitude do IC)=Mean-(0.178010-0.165099)=0.190922
```

b) moda =0; mediana =0 e média = 0.17801 ovos/folha

A média é superior à moda=mediana, devido à assimetria positiva que os dados apresentam Observe-e tb por exemplo o coeficiente de assimetria, *Skewness* =5.136906. Há um excesso de zeros e depois alguns valores são elevados, deslocando a média para a direita.

c) Para construir o *boxplot* devem calcular-se as barreiras superiores e inferiores, para se fazer uma pesquisa de candidatos a *outliers.*

$BI = Q_1 - 1.5(Q_3 - Q_1)$ $\quad BS = Q_3 + 1.5(Q_3 - Q_1)$, como $Q_1 = 0$, $Q_3 = 0$ tem-se $BI = 0$ e $BS = 0$, portanto não há valores observados inferiores à barreira inferior (logo não há outliers na cauda esquerda), mas os valores 1, 2, 3, 4, 5, 6, 7, 8 e 9 são todos eles candidatos a outliers. Ver abaixo o *boxplot* e por curiosidade o histograma (não era pedido),



d) Uma estimativa da proporção de folhas com ovos é dada por $p^{\star} = xtot/7640 = 0.125$ (está também no final de qualquer dos comandos `binom (...)`, como

```
sample estimates:
probability of success
                 0.125
```

e) Um intervalo de confiança a 95% para a proporção de folhas com ovos está no *output*

```
95 percent confidence interval:
 0.1176628 0.1326246
```

$0.1176628 < p < 0.1326246$, significando que, com 95% de confiança a verdadeira proporção de folhas com ovos está entre aqueles dois valores.

O intervalo acima é exacto (no sentido de ser usada a distribuição binomial na sua construção). Podemos usar a expressão do IC assintótico (que aqui não vai dar um resultado muito diferente, visto n ser muito grande)

$\hat{p}-z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p}+z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} \Leftrightarrow 0.125-1.96\sqrt{\frac{0.125\times 0.875}{7640}} < p < 0.125+1.96\sqrt{\frac{0.125\times 0.875}{7640}}$
donde o IC a 95% para p é $\quad]0.11758\,,\,0.1324[$

f) A resposta a esta questão deve ser dada construindo um teste de hipóteses:
$H_0 : p = 0.10 \qquad vs \qquad H_1 : p > 0.10,$ vamos considerar o nível de significância $\alpha = 0.05$

Como $n = 7640$ permite considerar a aproximação pela normal, sendo a Estatística de Teste
$Z = \dfrac{X - n\ p_0}{\sqrt{np_0q_0}} \sim \mathcal{N}(0,1)$

A região de Rejeição ou região Crítica é $RC : Z > z_\alpha \Leftrightarrow Z > 1.65$

Calculando o valor da Estatística de teste, no caso de H_0 verdadeira, i.e. quando $p \equiv p_0 = 0.1$, $Z_{calc} = \dfrac{955-746}{\sqrt{7640\times 0.1\times 0.9}} = 7.97$, portanto $Z_{calc} \in RC$ logo rejeitamos H_0, o que significa que, pode considerar-se a percentagem de folhas com ovos superior a 10%, com um nível de significância de 5%.

Nota: A resposta podia ser dada usando o *output*. O comando a escolher seria
`> binom.test(x=xtot, n=7640, p = 0.1,alternative = "greater")`
que realiza um teste exacto, mas para o qual se teve `p-value = 1.056e-12`, que portanto conduz à rejeição da hipótese nula.

g) Um intervalo de confiança a 95% para o número médio de pinheiros afectados/parcela é dado tb no output $\quad 1.819692 < \mu < 2.130308$

2. a) Temos $\theta > 0$ e o método dos momentos estabelece que um estimador é a solução de

$$E[X] = \frac{\sum X_i}{n} \Longleftrightarrow \theta\sqrt{\pi/2} = \overline{X} \Longleftrightarrow \theta = \overline{X}\sqrt{2/\pi}$$

Então o estimador é $\Theta^* = \overline{X}\sqrt{2/\pi}$

b) Um estimador T de θ é centrado se e só se $E[T] = \theta$.
Vamos então calcular $E[\Theta^*] = E[\overline{X}\sqrt{2/\pi}] = \sqrt{2/\pi}\,E[\overline{X}]$, pelas propriedades do valor médio. Como $E[\overline{X}] = E[X] = \theta\sqrt{\pi/2}$, substituindo em cima dá $E[\Theta^*] = \theta$, portanto é estimador centrado.

c) Para obter o Estimador de Máxima Verosimilhança é preciso escrever a função de verosimilhança:

$$L(\theta|x_1,\ldots,x_n) = \frac{\Pi_{i=1}^n x_i}{\theta^{2n}} exp\left(-\frac{\sum_{i=1}^n x_i^2}{2\theta^2}\right)$$

O logaritmo é

$$logL(\theta|x_1,\ldots,x_n) = log\left(\Pi_{i=1}^n x_i\right) - 2\ n\ log(\theta) - \frac{\sum_{i=1}^n x_i^2}{2\theta^2}$$

Calculando a derivada:

$$\frac{dlogL}{d\theta} = 0 - \frac{2n}{\theta} - \frac{1}{2}\frac{(-2\theta)\sum_{i=1}^n x_i^2}{\theta^4} = -\frac{2n}{\theta} + \frac{\sum_{i=1}^n x_i^2}{\theta^3}$$

e agora igualando a zero:

$$-\frac{2n}{\theta} + \frac{\sum_{i=1}^n x_i^2}{\theta^3} = 0 \Longleftrightarrow -2n + \frac{\sum x_i^2}{\theta^2} = 0 \Longrightarrow \frac{\sum x_i^2}{\theta^2} = 2n \Longleftrightarrow \theta^2 = \frac{\sum x_i^2}{2n}$$

O estimador de máxima verosiminhança é $\widehat{\Theta} = \sqrt{\dfrac{\sum X_i^2}{2n}}$

d) Usando os dois estimadores obtidos na alínea anterior podemos ter 2 estimativas de θ com base na amostra observada.

$\theta^* = \overline{x}\sqrt{2/\pi} = 2.2378$ e $\widehat{\theta} = \sqrt{\dfrac{\sum x_i^2}{2n}} = 2.114496$

e) $CV = \dfrac{\sigma}{\mu} \times 100\%$. Para estimar o CV podemos usar directamente as estimativas de σ e μ, que são s e $\overline{x}$ ou então usar uma estimativa de θ e depois substituir na expressão da variância (a raíz quadrada é o desvio padrão estimado) e do valor médio.

3. a) Uma estimativa para a diferença média entre os valores de pH da água à entrada e à saída do tanque, obtém-se facilmente do *output* em

```
sample estimates:
mean of the differences
-0.3466667
```

b) Para averiguarmos se há alteração média diária do valor do pH, vamos realizar um teste de hipóteses à diferença de médias.

Como as observações são registadas dia a dia as amostras recolhidas devem ser tratatadas como emparelhadas por dia.

Como temos uma amostra (emparelhada) de dimensão pequena ($n = 15$), para a realização de um teste$-t$ necessitamos de primeiro testar a hipótese da normalidade, analisando o resultado de

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

onde se obteve um `p-value = 0.8509`, levando portanto à não-rejeição da hipótese da normalidade. O teste t a considerar está no comando

```
> 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
```

para o qual se obteve `p-value = 0.002540`, conduzindo portanto à rejeição da hipótese nula ($H_o : \mu_1 = \mu_2$) e a considerar-se existir diferença significativa entre as médias. Então há alteração média diária do valor do pH quando a água passa no tanque, mais concretamente o valor médio do pH aumenta.

4. Se X e Y são variáveis aleatórias **contínuas** com distribuição uniforme em $[0, 1]$:

a) `x<-runif(1000)`
`y<-runif(1000)`

b) `z<-x+y`

c) `hist(z)`

d) Da leitura do comando

```
> 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
```

Aplicou-se o teste de ajustamento de Kolmogorov-Smirnov, com hipótese nula uniforme, à amostra resultante da soma de 2 variáveis uniformes, somos levados à rejeição da hipótese de que a soma é uniforme.

```
> 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
```

Aplicou-se o teste de ajustamento de Kolmogorov-Smirnov, com hipótese nula normal, à amostra resultante da soma de 2 variáveis uniformes, somos levados à não rejeição da hipótese de que a soma segue uma lei normal.

5. Se $(X_1, X_2, \ldots, X_n)$ tal que $X \frown \mathcal{N}(\mu_1, \sigma) \Longrightarrow \overline{X} \frown \mathcal{N}(\mu_1, \sigma/\sqrt{n})$

$(Y_1, Y_2, \ldots, Y_n)$ tal que $Y \frown \mathcal{N}(\mu_2, \sigma) \Longrightarrow \overline{Y} \frown \mathcal{N}(\mu_2, \sigma/\sqrt{n})$.

amostras independentes e σ conhecido.

a) As hipóteses a testar são: $H_0 : \mu_1 = \mu_2 \quad vs \quad H_1 : \mu_1 \neq \mu_2$, portanto a estatística de teste será:
$\overline{X} - \overline{Y} \frown \mathcal{N}(\mu_1 - \mu_2, \sqrt{\sigma^2/n + \sigma^2/n}) \Longleftrightarrow \overline{X} - \overline{Y} \frown \mathcal{N}(\mu_1 - \mu_2, \sigma\sqrt{2/n})$.
Como um IC para μ_1 é da forma $\overline{x_1} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ e um IC para μ_2 é da forma $\overline{x_2} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$, tal como está na Nota, eles não se sobrepõem se a diferença entre as médias amostrais for superior à soma das semi-amplitudes de cada um dos intervalos (cada uma dada por $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$).
Logo sobrepõem-se quando aquela diferença for inferior à soma das semi-amplitudes, ou seja for inferior a $2\ z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

P[os intervalos se sobreporem | H_0 verdadeira] =
$P[|\overline{X} - \overline{Y}| < 2\ z_{\alpha/2}\frac{\sigma}{\sqrt{n}}] = P[-2\ z_{\alpha/2}\ \frac{\sigma}{\sqrt{n}} < \overline{X} - \overline{Y} < 2\ z_{\alpha/2}\ \frac{\sigma}{\sqrt{n}}] =$
$$= P\left[\frac{-2\ z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}{\sigma\sqrt{2/n}} < Z < \frac{2\ z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}{\sigma\sqrt{2/n}}\right] = P[-\sqrt{2}z_{\alpha/2} < Z < \sqrt{2}\ z_{\alpha/2}] = \Phi(\sqrt{2}\ z_{\alpha/2}) - \Phi(-\sqrt{2}\ z_{\alpha/2}) =$$
$2\Phi(\sqrt{2}\ z_{\alpha/2}) - 1$

b) Basta substituir para $\alpha = 0.05$, dando $2\Phi(\sqrt{2} \times 1.96) - 1 = 0.994$.

c) P[erro tipo I]=P[rejeitar H_0 |H_0 verdadeira]=
P[intervalos não se sobreporem |H_0 verdadeira]=0.05
então tem-se
$1 - 2\Phi(\sqrt{2}\ z_{\alpha/2}) + 1 = 0.05 \Longleftrightarrow \Phi(\sqrt{2}\ z_{\alpha/2}) = 0.975 \Longleftrightarrow \sqrt{2}\ z_{\alpha/2} = 1.96 \Longleftrightarrow \alpha = 0.16577$