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

Teste do Módulo I (2ª data) – duração: 2h30 27 de Junho de 2017

Nota: Alguns cálculos de apoio à resolução deste teste encontram-se em Anexo

1. Contou-se o número de folhas em cada uma de 150 plantas de tabaco. Os resultados encontram-se na seguinte tabela: ‘

Número de Folhas	17	18	19	20	21	22	23	24
Número de Plantas	3	22	44	42	22	10	6	1

a) Qual é a variável em estudo? Classifique-a, justificando a sua resposta.

b) Calcule a média, a mediana e a variância da variável indicada em a).

c) Desenhe o *boxplot* dos dados registados e comente a distribuição dos dados.

d) Calcule, interpretando-o, um intervalo a 95% de confiança para o valor médio do número de folhas por planta de tabaco.

e) Comente, justificando adequadamente, a seguinte afirmação: “Pelo menos 25% das plantas têm mais de 20 folhas”.

2. A distribuição de Gumbel é uma das distribuições que permite modelar o comportamento do máximo de séries de observações, X, sob certas condições. Este modelo, na sua forma completa, depende de dois parâmetros desconhecidos, mas aqui vamos considerar apenas desconhecido o parâmetro de escala $\delta > 0$, i.e., a função densidade de probabilidade da varável em estudo é:

$$f(x|\delta) = \frac{1}{\delta}\, exp\left(-\frac{x}{\delta}\right) \;\; exp\left[-exp\left(-\frac{x}{\delta}\right)\right], \qquad x \in \mathbb{R},$$

Sabe-se que $E[X] = \gamma\delta$, com $\gamma \simeq 0.577215$, *constante de Euler*. Considere que dispõe de uma amostra aleatória de dimensão n, $(X_1, X_2, \ldots, X_n)$, associada a X.

a) Determine o estimador de δ calculado pelo método dos momentos.

b) Mostre que uma estimativa de máxima verosimilhança para se obtém como solução da equação:

$$\overline{x} = \delta + \frac{\sum x_i \; exp\left(-\frac{x}{\delta}\right)}{n}.$$

c) Como pode verificar a equação acima não permite explicitar o valor da estimativa $\widehat{\delta}$. Consulte o **Anexo**, onde se encontra uma amostra observada, que foi introduzida no R, apresentam-se vários comandos e os respectivos resultados. Responda às seguintes questões:

i) Explique sucintamente o que foi feito no conjunto de comandos identificado como ““**Grupo 1 de comandos”**.

ii) O que lhe permite concluir o gráfico apresentado?

iii) Indique duas estimativas de δ .

3. O método habitual de análise de solos para detecção de nemátodes, por decantação e crivação, foi aplicado a amostras de vários solos. Foi efectuada a contagem do número de nemátodes em 39 amostras (de 300 g cada) desses solos, os dados foram introduzidos no R no vector `DeCriv` e executaram-se alguns comandos, tendo-se obtido os seguintes resultados:

No laboratório de nematologia foi implementado o método de análise de solos para detecção de nemátodes recomendado internacionalmente, o método Fenwick, que se supõe ser mais eficiente que o método habitual, embora este seja rápido e barato.

Às mesmas amostras de solo foi aplicado este método e as contagens obtidas foram introduzidas no R no vector `Fen`. Apresentam-se no **Anexo** alguns dos comandos e respectivos resultados.

Com base nos valores obtidos poder-se-á concluir que o número de nemátodes que foram detectados pelo método Fenwick é, em média, significativamente superior ao detectado pelo método decantação+crivação? Responda justificando convenientemente.

4. Os valores de pH obtidos em 16 análises de água mineral de uma certa origem são:

6.7, 6.1, 5.9, 6.9, 7.6, 7.4, 7.4, 7.2, 6.3, 5.6, 5.9, 7.0, 7.8, 6.7, 6.0, 7.1

a) Será de admitir que valor do pH das águas analisadas provenientes dessa origem segue uma lei normal? Justifique.

b) Obtenha um intervalo de confiança a 99% para o verdadeiro valor médio do pH. Comente.

c) Poderemos concluir, ao nível de significância de 5%, que o pH médio da água proveniente da captação é superior a 6.5?

d) Suponha que alguém questionou as conclusões obtidas, alegando que a metodologia usadas nas análises não tinha sido a adequada. Realizaram-se de novo 16 análises de água mineral da mesma origem, tendo-se obtido os seguintes resultados:

6.7, 6.6, 5.8, 5.2, 5.7, 7.1, 7.9, 6.4, 6.5, 5.6, 6.0, 6.6, 6.4, 6.7, 5.4, 6.0

i) Haverá efectivamente diferença nas metodologias de análise utilizadas? Justifique convenientemente, indicando e validando os pressupostos necessários à resolução da questão.

ii) Usando esta amostra manter-se-ia a resposta que deu em c)?

5. Considere X uma população $N(\mu, \sigma)$, com parâmetros desconhecidos. Seja $(X_1, \ldots, X_n)$ uma amostra aleatória retirada de X.

a) Considere $S^2 = \dfrac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1} = \dfrac{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}{n-1}$. Mostre que S^2 é um estimador centrado de σ^2.

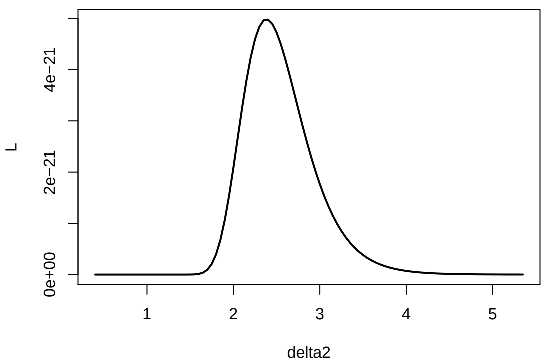
b) Considere um outro estimador de σ^2, $\quad V^2 = \dfrac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}$.

i) Mostre que $E[V^2] = \dfrac{n-1}{n}\sigma^2$ e calcule o viés de V^2, $Bias[V^2]$. Comente.

ii) Sabendo que $Var[S^2] = \dfrac{2\sigma^4}{n-1}$, obtenha o *Erro Quadrático Médio* (EQM) de cada um dos estimadores S^2 e V^2.

ANEXO

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


> plantas<-c(3,22,44,42,22,10,6,1)
> folhas<-c(17,18,19,20,21,22,23,24)
> plantas<-c(3,22,44,42,22,10,6,1)
>
> sum(folhas^2*plantas)
[1] 58971


########  Pergunta 2   ########

### Grupo 1 de comandos

> x<-c(-1.13, -4.13, -0.38,  1.67,  0.27,  1.91,  4.51,  0.56, -1.94, -0.89, -0.67,  1.59,
+  1.24,  4.26, -0.26,  2.74, -0.39, -2.94,  4.87,  1.22)

> library(evd)

> L<-c();delta2<-c()
> ini<-0.35
> for (i in 1:100)  {
+    delta2[i]<-ini+0.05
+    L[i]<-prod(dgumbel(x, scale=delta2[i]))
+    ini<-delta2[i]
+ }
> plot(delta2,L,type="l")
###  fim do Grupo 1
```




```
> logL_negativa <- function(delta){
+   -sum(dgumbel(x, scale=delta, log = TRUE))}
>
> delta.est<-mean(x)/0.5772
> delta.est
[1] 1.049030
> library(stats4)
> MLest <- mle(logL_negativa, start = list(delta=delta.est))
> summary(MLest)
Maximum likelihood estimation

Call:
mle(minuslogl = logL_negativa, start = list(delta = delta.est))

Coefficients:
      Estimate Std. Error
```

```
delta 2.381733  0.3449365

-2 log L: 93.49582


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


> Fen<-c(0,2,8,4,2,2,1,2,13,15,19,2, ... ,2,1,3,0,60,7,1,27,204)
> DeCriv<-c(1,5,2,1,2,0,0,1,2,3,2,2, ... ,0,0,18,3,17,0,0,0,47)

> shapiro.test(Fen)

  Shapiro-Wilk normality test

data:  Fen
W = 0.5213, p-value = 4.198e-10

> shapiro.test(DeCriv)

  Shapiro-Wilk normality test

data:  DeCriv
W = 0.5989, p-value = 4.069e-09

> shapiro.test(Fen-DeCriv)

        Shapiro-Wilk normality test

data:  Fen - DeCriv
W = 0.618, p-value = 7.42e-09

> mean(Fen);mean(DeCriv)
[1] 18.38462
[1] 8.179487

> var(Fen);var(DeCriv);var(Fen-DeCriv)
[1] 1408.769
[1] 220.9933

[1] 812.8516

########  Pergunta 4   ########



> x<-c(6.7,6.1,5.9,6.9,7.6,7.4,7.4,7.2,6.3,5.6,5.9,7.0,7.8,6.7,6.0,7.1)
> y<-c(6.7,6.6,5.8,5.2,5.7,7.1,7.9,6.4,6.5,5.6,6.0,6.6,6.4,6.7,5.4,6.0)
> shapiro.test(x);shapiro.test(y);shapiro.test(x-y)

        Shapiro-Wilk normality test

data:  x
W = 0.9487, p-value = 0.4701


        Shapiro-Wilk normality test

data:  y
W = 0.959, p-value = 0.6442


        Shapiro-Wilk normality test

data:  x - y
W = 0.9162, p-value = 0.1468

>
> t.test(x)

        One Sample t-test

data:  x
t = 39.2935, df = 15, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
```

```
95 percent confidence interval:
 6.360206 7.089794
sample estimates:
mean of x
    6.725

> t.test(x,mu=6.5)

        One Sample t-test

data:  x
t = 1.3147, df = 15, p-value = 0.2084
alternative hypothesis: true mean is not equal to 6.5
95 percent confidence interval:
 6.360206 7.089794
sample estimates:
mean of x
    6.725

> t.test(x,conf.level = 0.99)

        One Sample t-test

data:  x
t = 39.2935, df = 15, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
99 percent confidence interval:
 6.220676 7.229324
sample estimates:
mean of x
    6.725

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

        One Sample t-test

data:  x
t = 39.2935, df = 15, p-value < 2.2e-16
alternative hypothesis: true mean is greater than 0
95 percent confidence interval:
 6.424969      Inf
sample estimates:
mean of x
    6.725

> t.test(x,alternative = "greater",mu=6.5)

        One Sample t-test

data:  x
t = 1.3147, df = 15, p-value = 0.1042
alternative hypothesis: true mean is greater than 6.5
95 percent confidence interval:
 6.424969      Inf
sample estimates:
mean of x
    6.725
```

```
> t.test(x,y)

        Welch Two Sample t-test

data:  x and y
t = 1.8058, df = 30, p-value = 0.081
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.05729476  0.93229476
sample estimates:
mean of x mean of y
   6.7250    6.2875

> t.test(x,y,var.equal=T)

        Two Sample t-test

data:  x and y
t = 1.8058, df = 30, p-value = 0.081
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.05729468  0.93229468
sample estimates:
mean of x mean of y
   6.7250    6.2875

> t.test(x,y,paired=T)

        Paired t-test

data:  x and y
t = 2.3361, df = 15, p-value = 0.03378
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.03832715 0.83667285
sample estimates:
mean of the differences
                 0.4375

> t.test(x,y,alternative="greater",var.equal=T)

        Two Sample t-test

data:  x and y
t = 1.8058, df = 30, p-value = 0.0405
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.02629352        Inf
sample estimates:
mean of x mean of y
   6.7250    6.2875
```