Exercícios a enviarem a Profª Manuela Neves (manela@isa.ulisboa.pt) até ao final do dia 7 de Maio

Nota: Os exercícios 1 e 2 referem-se ao 1º bloco e os exercícios 3 e 4 referem-se ao 2º bloco de aulas teóticas

1. [*1º Teste de 2018/2019*] Foi apresentado na *Nature* um estudo realizado em Inglaterra e País de Gales em que se pretendia averiguar se, nos anos seguintes a períodos de guerra, o sexo do primeiro filho estava relacionado com a diferença de idade entre os pais (idade do pai ? idade da mãe). O estudo foi feito numa amostra de famílias tendo-se registado as idades dos pais e o sexo do primeiro filho. O dados encontram-se na seguinte tabela:

Sexo do 1º filho	Idade do pai-idade da mãe		
	-9 a -1	0 a 5	5 a 15
Masculino	14	117	37
Feminino	29	84	20

Consulte o **Anexo** abaixo com alguns resultados obtidos no R para responder, quando necessário, às seguintes perguntas:

a) Qual a dimensão da amostra recolhida?

b) Complete os valores e expressões em falta que foram substituídos pelas letras **A**, **B**, **C**, **D**, **E**, **F**.

c) Poder-se-á considerar haver relação entre a diferença de idades dos pais e o sexo do primeiro filho? Justifique convenientemente.

d) Interprete os resíduos do teste, para complementar as conclusões a que chegou em iii).

ANEXO

```
> estudo.sexo<-matrix(c(14,117,37,29,84,20),nc=3,byrow = A ,
  +   dimnames=list(c("Masculino", "Feminino"),c("-9 a -1 ",
+   " 0 a 5"," 5 a 15")))

> estudo.sexo
          -9 a -1    0 a 5   5 a 15
Masculino       14     117       37
Feminino        29      84       20

> margin.table(B,1)                    > margin.table(B,2)
Masculino  Feminino                   -9 a -1     0 a 5     5 a 15
      168       133                       43        201        57


> teste<-chisq.test(estudo.sexo)
> teste

        Pearson's Chi-squared test
```

```
data:  estudo.sexo
X-squared = 11.8106, df = C, p-value = 0.002725

> teste$expected
            -9 a -1       0 a 5     5 a 15
Masculino       24   112.18605     31.81395
Feminino         D   88.81395     25.18605


> teste$residuals
           -9 a -1       0 a 5     5 a 15
Masculino -2.041241   0.4544985    0.9194489
Feminino   2.294157  -0.5108122     E

> pchisq(teste$statistic,2)
  F
```

2. [*1º Teste de 2018/2019*] Considere os seguintes comandos e o output associado:

```
> dna<-c("A","C","G","T")
   >  seq2<-sample(dna,1000,replace=T,
  +  prob=c(0.20,0.30,0.18,0.32))

> table(seq2)
 seq2
   A    C    G    T
  192  289  183  336

> pbinom(192,1000,0.20)
[1] 0.2783474

> 1-pbinom(207,1000,0.20)
[1] 0.2749125
```

Explique o que cada linha de comandos executa e diga se é admissível supor que o nucleótido "A" ocorre na proporção definida no estudo.

3. [*2º Teste de 2017/2018*] Suponha que dispõe de uma amostra aleatória de dimensão n, $(X_1, X_2, .., X_n)$, retirada de uma população X, com função densidade definida abaixo, onde $\beta > -1$ é um parâmetro desconhecido:

$$f(x;\beta) = \begin{cases} \dfrac{\beta+1}{\mathrm{e}^{\beta+1}} x^{\beta} & \text{se } \ 0 \leq x \leq \mathrm{e} \\ 0 & \text{outros valores de } x \end{cases}$$

Nota: Sabe-se que $E[X] = \dfrac{(\beta+1)\mathrm{e}}{\beta+2}$.

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

b) Obtenha o estimador de máxima verosimilhança para β.

c) Considere os valores observados da seguinte amostra, de dimensão 30, extraída daquela população, com a qual se realizaram os cálculos apresentados:

```
> dados
 [1] 2.13 2.26 1.32 1.88 1.63 2.27 2.65 2.60 1.80 2.00
[11] 1.85 1.54 0.63 2.47 1.67 2.14 1.47 2.30 2.02 2.57
[21] 1.77 2.40 1.96 2.33 2.12 1.62 0.82 2.39 2.57 2.59

> sum(dados)              > sum(log(dados))
[1] 59.77                  [1] 19.39522
```

d) Determine estimativas para β.

4. [*2º Teste de 2017/2018*] Considere o seguinte estimador do parâmetro β, do modelo tratado na pergunta anterior, assim definido:

$$\beta^* = \frac{2(\overline{X}-1)}{2-\overline{X}}.$$

A obtenção do valor médio e da variância deste estimador não é fácil com recurso a procedimentos de Estatística Clássica, pelo que se considerou o recurso à metodologia *bootstrap*.

a) Complete o que deve estar nas letras em falta indicadas no *output* por A, B e C

b) Determine uma estimativa *bootstrap* de β.

c) Obtenha um intervalo *booststrap* a 90% de confiança para β.

ANEXO

```
> n<-length(dados);n
[1] A
> beta.est<-vector()
> for(i in 1:1000)
+  {   dados_star<-sample(dados,n,rep=B)
+  beta.est[i]<-2*(mean(dados_star)-1)/(3-mean(dados_star))  }
> sum(beta.est)
[1] 2003.918
> mean(beta.est)
[1] C
>
> quantile(beta.est,prob=c(0.01,0.025,0.05,0.10,0.90,0.95,0.975,0.99))
      1%      2.5%        5%       10%       90%       95%     97.5%       99%
1.275056  1.376359  1.465054  1.585193  2.467943 2.588998  2.792428  2.948597
```