Instituto Superior de Agronomia
Modelos Matemáticos e Aplicações – 2017/2018

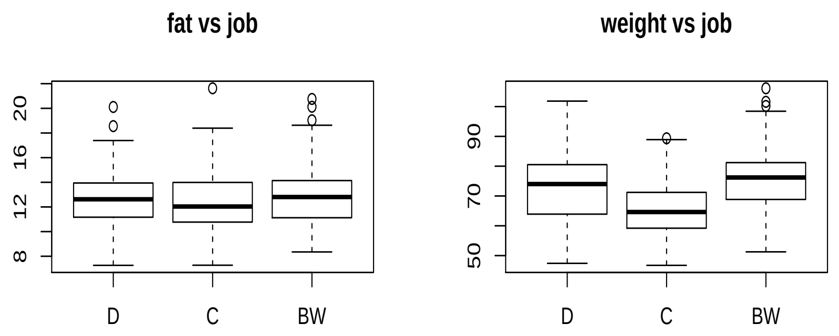
Teste do Módulo I– duração: 2h30 14 de Março de 2018

1. Considere os dados `diet` contidos na biblioteca `Epi` do R, referentes a um estudo que tinha como objectivo analisar a existência de associação entre a dieta e o aparecimento de doença coronária. A *data frame* `diet` tem 337 linhas e 15 colunas. Vamos considerar apenas as seguintes variáveis:

`fat`: fat intake (g/day), um vector numérico
`weight`: (kg), um vector numérico
`chd`: CHD event, um vector tomando valores (1=CHD event, 0=no event)
`job`: ocupação, um factor com os níveis `Driver`(D), `Conductor`(C) e `Bank worker`(BW)

Foram efectuados alguns comandos 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** e **D**, em falta no *output*.
b) Esboce o **Histograma**, que aparece identificado no *output*, mas não foi desenhado.
c) Faça o esboço de um *boxplot* para a variável `fat`. Indique claramente os limites do *boxplot*.
d) Comente, comparando, os 2 grupos de *boxplots* apresentados abaixo.
e) Obtenha um intervalo de confiança a 95% para a proporção de indivíduos em que **não terá havido ocorrência de doença coronária**. Comente.
f) Face aos resultados apresentados no *output* compare o valor médio da variável `fat` entre cada um dos grupos de classificação de ocorrência ou não de doença coronária. Justifique.



2. Um agrónomo está a desenvolver uma variedade melhorada de pimenta verde. Os supermercados indicaram que os clientes geralmente não compram a vagens de pimenta verde que pesam menos de 45 g. A actual variedade de planta produz pimentas verdes que pesam 48 gramas em média, mas 13% pesam menos de 45 gramas. Suponha que o peso da vagem da variedade actual de pimentas verdes segue uma distribuição normal.

a) Qual é o desvio padrão do peso da variedade actual de pimentas verdes?
b) O agrónomo pretende reduzir a frequência de vagens de pimenta verde que pesam menos de 45 gramas para 5% no máximo. Uma maneira de reduzir a frequência de pimentas verdes com baixo peso é tentar aumentar o peso das pimentas verdes. Se o desvio padrão permanecer o mesmo, qual o valor médio que deve ser procurado atingir? (se não resolveu a) considere $\sigma = 2.6$).
c) O agrónomo produziu a nova variedade de pimentas verdes, por forma a garantir um peso médio de 50 gramas com um desvio padrão de 2.5 gramas. Um cliente do supermercado compra 25. Qual a probabilidade de a média do peso dessas pimentas ser inferior a 49 gramas?

3. Suponha que o número de tentativas que um jogador de basquetebol faz até acertar no cesto é uma variável aleatória, X, com a seguinte função massa de probabilidade:

$$P[X = x|\theta) = \frac{1}{\theta}\left(1 - \frac{1}{\theta}\right)^{x-1}, \quad \text{para} \;\; x = 1, 2, 3, ... \quad (\theta > 1)$$

Nota: Sabe-se que $E[X] = \theta$ e $Var[X] = \theta(\theta - 1)$.

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

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

b) Diga se o estimador dos momentos obtido na alínea a) é um estimador centrado de θ e indique o seu Erro Quadrático Médio.

c) Obtenha o estimador de máxima verosimilhança para θ.

d) Considere os valores observados da seguinte amostra, de dimensão 20, extraída daquela população, com a qual se realizaram os cálculos apresentados:

```
> amostra
 [1] 2  7  1  1  1  3  5  6  4  5  3  2  4  6  6  6  1  7  5  1
> sum(amostra)                   > sum(amostra*amostra)
[1] 76                           [1] 380
```

i) Determine uma estimativa para θ.

ii) Determine uma estimativa de $P[X = 2]$, i.e., da probabilidade de o jogador acertar à segunda tentativa.
Nota: o estimador de máxima verosimilhança goza de uma propriedade, dita da invariância, que estabelece que –*O estimador de máxima verosimilhança de $g(\theta)$ é $g(\hat{\Theta})$, se $g(.)$ é função biunívoca de θ.*

4. Indique, justificando se as seguintes afirmações são **Verdadeiras** ou **Falsas**. **Corrija as falsas**.

a) Se $X \frown Poisson(10)$ então pode calcular-se $P[X > 8]$, com recurso ao R fazendo `dpois(8,10)`.

b) Suponha que $Y \frown Binomial(n, p)$. Então $P[0 \leq Y \leq n] = 1$.

c) Seja $X \frown N(1, 2)$ e $Y \frown N(1, 1)$, com X e Y independentes. Então $2X - Y \frown N(1, 3)$.

d) Seja $(X_1, ..., X_n)$ uma amostra aleatória de tamanho n, proveniente de uma população com valor médio μ e variância $\sigma^2 < +\infty$. Então para n suficientemente elevado tem-se $P[S_n \leq n\mu] \approx 1/2$, onde $S_n = \sum_{i=1}^{n} X_i$.

5. Admite-se que a concentração, X, de dióxido de enxofre (SO_2) (em partes por milhão, ou ppm) numa determinada cidade segue uma distribuição gama, que vimos poder ter dois parâmetros α e β, ($\alpha > 0$, $\beta > 0$). Vamos considerar, para simplificar, que $\alpha = 1/2$, permanecendo apenas β desconhecido. A função densidade é então assim dada:

$$f(x|\beta) = \begin{cases} \dfrac{1}{\sqrt{\beta\pi x}}\, e^{-x/\beta} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Seja $(X_1, ..., X_n)$ uma amostra aleatória de tamanho n, proveniente daquela lei, sendo $(x_1, ...x_n)$ a correspondente amostra observada.

a) Usando o Teorema Limite Central obtenha uma distribuição aproximada para $\overline{X}$.

b) Recordando que se uma v.a. V é tal que $V \sim Normal(0, 1)$, se pode escrever

$$P[-1.96 < V < 1.96] \approx 0.95,$$

obtenha um intervalo assintótico a 95% de confiança para o parâmetro β.

ANEXO I

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

> library(Epi)
> data(diet)
> dim(diet)
[1] 337  15

> names(diet)
 [1] "id"         "doe"        "dox"        "dob"        "y"
 [6] "fail"       "job"        "month"      "energy"     "height"
[11] "weight"     "fat"        "fibre"      "energy.grp" "chd"


> hist(diet$fat,breaks=c(5,10,12,15,18,22),plot=F) #Histograma
$breaks
[1]  5 10 12 15 18 22

$counts
[1]  34 106 146  41  10

$density
[1] 0.020178042 0.157270030 0.144411474 0.040553907 0.007418398

$mids
[1]  A  11.0  13.5  16.5  20.0

>
> sort(diet$fat)
[1]  7.260000  7.272000  7.578000  8.302000  8.348001  8.536000  8.577000
. . .
[323] 17.388000 17.409000 17.445000 17.639000 17.769000 18.389000 18.391000
[330] 18.558000 18.588000 18.628000 19.032000 20.112000 20.132000 20.763000
[337] 21.629000


> summary(diet$fat)
   Min.  1st Qu.   Median    Mean   3rd Qu.    Max.
   B     11.12     12.59     12.75   14.01     21.63


> par(mfrow=c(1,2))
> boxplot(fat~job,data=diet,main="fat vs job",names=c("D","C","BW"))
> boxplot(weight~job,data=diet,main="weight vs job",names=c("D","C","BW"))


> diet$fat[diet$chd==0]    # C - escreva, brevemente, o que resulta deste comando
> length(diet$fat[diet$chd==0])
[1] 291
> length(diet$fat[diet$chd==1])
[1] D
```

```
> shapiro.test(diet$fat[diet$chd==0])

        Shapiro-Wilk normality test

data:  diet$fat[diet$chd == 0]
W = 0.9757, p-value = 7.604e-05


> shapiro.test(diet$fat[diet$chd==1])

        Shapiro-Wilk normality test

data:  diet$fat[diet$chd == 1]
W = 0.9743, p-value = 0.3947


> t.test(fat~chd,data=diet)

        Welch Two Sample t-test

data:  fat by chd
t = 2.9536, df = 62.484, p-value = 0.004423
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.337497    1.750317
sample estimates:
mean in group 0 mean in group 1
       12.88791        11.84400


> t.test(fat~chd,data=diet,alternative="greater")

        Welch Two Sample t-test

data:  fat by chd
t = 2.9536, df = 62.484, p-value = 0.002211
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.4537988       Inf
sample estimates:
mean in group 0 mean in group 1
       12.88791        11.84400



> t.test(fat~chd,data=diet,alternative="less")

        Welch Two Sample t-test

data:  fat by chd
t = 2.9536, df = 62.484, p-value = 0.9978
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
     -Inf 1.634016
sample estimates:
mean in group 0 mean in group 1
       12.88791        11.84400

>
```