Instituto Superior de Agronomia
Estatística – 1ª Chamada (2020/2021)

Duração: 3 horas 7 de Janeiro de 2021

1. (**4 val.**) Num estudo sobre a evolução temporal da superfície ocupada com as principais culturas agrícolas de Portugal, registou-se durante 34 anos a área de cereais para grão (em km^2).

a) Os valores dessa área foram introduzidos no R no vector `area.cereal` e obtidos os resultados dos seguintes comandos:

```
> min(area.cereal)                    > max(area.cereal)
[1] 2243                              [1] 9075

> hist(area.cereal,breaks=c(2200,3200,4200,6200,8200,9200),plot=FALSE)
$breaks
[1] 2200 3200 4200 6200 8200 9200
$counts
[1] 11  4  7  7  5
$density
[1] 0.0003235294 0.0001176471 0.0001029412 0.0001029412 0.0001470588
$mids
[1] 2700 3700 5200 7200 8700
```

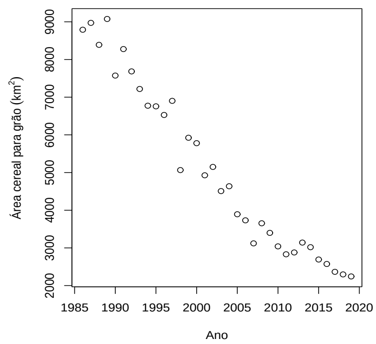
i) No comando `hist()` há um argumento que impediu o R de desenhar o histograma. Qual foi? Esboce o histograma que o R desenharia.

ii) Determine o valor aproximado da média e da mediana da área de cereais para grão.

b) Para os dados recolhidos, foi construído o diagrama de dispersão entre o ano (variável x) e a área de cereais para grão (variável y, em km^2), no período em estudo: de 1986 a 2019 (figura ao lado). Foi ajustada a recta de regressão dos mínimos quadrados de y sobre x, cuja equação é

$$y = 444750.1 - 219.5\ x,$$

com uma precisão de 95.28%.



i) Determine, justificando, o coeficiente de correlação entre x e y.

ii) Indique, justificando, se considera adequada a utilização da recta de regressão para descrever a evolução temporal da área de cereais para grão entre 1986 a 2019.

iii) Determine o resíduo associado ao par observado (2000, 5779).

iv) Determine os valores dos coeficientes da nova recta de regressão se a área de cereais para grão fosse expressa em hectares ($1\ km^2 = 100$ ha).

2. (**3 val.**) Para poder ser contratado por uma certa empresa, um candidato tem que realizar duas provas, A e B, independentes. Cada prova tem a classificação de insuficiente (0), suficiente (1) ou bom (2). Sejam X e Y as variáveis aleatórias que representam as classificações das provas A e B, respectivamente, que se admitem ter as seguintes distribuições:

x_i	0	1	2
$P[X = x_i]$	0.2	0.5	0.3

y_i	0	1	2
$P[Y = y_i]$	0.2	0.6	0.2

a) Determine a função de distribuição cumulativa da v.a. X.

b) Determine a função massa de probabilidade conjunta das classificações obtidas nas duas provas.

c) Qual a probabilidade do candidato ter uma classificação igual nas duas provas?

d) Determine a função massa de probabilidade condicional de Y dado $X = 2$.

3. (**4 val.**) O número de clientes atendidos numa loja de telemóveis segue um processo de Poisson com taxa média de 8 clientes por hora.

a) Determine a probabilidade de numa hora serem atendidos:

i) no máximo 2 clientes;

ii) pelo menos 5 clientes.

b) Sempre que, numa hora, são atendidos pelo menos 5 clientes, a probabilidade de haver a compra de algum telemóvel é de 0.20; quando são atendidos no máximo 2 aquela probabilidade é 0.60; quando o número de clientes atendidos é 3 ou 4 aquela probabilidade é 0.50. Escolhida uma hora de atendimento ao acaso, qual a probabilidade de ter ocorrido a compra de algum telemóvel?

c) O horário dessa loja é das 16h às 20h todos os dias.

i) Qual o número médio de clientes atendidos por dia? Justifique.

ii) Numa amostra aleatória de trinta dias qual a probabilidade, aproximada, de a média diária de clientes atendidos ser superior a 35? Justifique.

4. (**3 val.**) Uma determinada vacina é distribuída em frascos que deverão ser suficientes para 5 doses de 0.3 mL, i.e, os frascos deverão conter no mínimo 1.5 mL. Uma máquina deita em cada frasco uma quantidade de vacina que se admite seguir uma lei normal com valor médio 1.75 mL e desvio padrão 0.1 mL.

a) Qual a probabilidade de um frasco escolhido ao acaso:

i) conter uma quantidade de vacina não suficiente para 5 doses?

ii) conter uma quantidade de vacina suficiente para 6 ou mais doses?

b) Os frascos são embalados em caixas que contêm 195 frascos. Qual é a probabilidade de numa caixa seleccionada ao acaso haver pelo menos um frasco que contém uma quantidade de vacina não suficiente para 5 doses? Justifique.

5. (**2.5 val.**) Seja (X, Y) um par de variáveis aleatórias contínuas, independentes.

a) Mostre que $E[XY] = E[X]E[Y]$.

b) Calcule, justificando, $Cov[X - 2, 2Y + 1]$.

c) Considere que X tem distribuição uniforme contínua em $]0, 1[$ e Y tem distribuição exponencial de valor médio 1.

Foram recolhidas as amostras aleatórias $(X_1, \ldots, X_n)$ e $(Y_1, \ldots, Y_n)$ de X e Y, respectivamente.

i) Defina "amostra aleatória de dimensão n extraída de X";

ii) Exprima, em função de n, $Var[\overline{X_n} - \overline{Y_n}]$, sendo $\overline{X_n}$ e $\overline{Y_n}$ as médias das amostras aleatórias $(X_1, \ldots, X_n)$ e $(Y_1, \ldots, Y_n)$, respectivamente. Justifique.

6. (**3.5 val.**) Um comprador afirma que o peso de pinhão (em gramas) que costuma obter ao comprar pinha de pinheiro manso é 35 gramas em média por pinha. Um proprietário, durante o período de negociação do preço, afirma que a sua pinha tem em média peso superior àquele valor. Por forma a demonstrar que tem razão propõe-se fazer uma amostragem. Como tem duas plantações de pinheiro manso em dois locais distintos, recolhe uma amostra aleatória de 49 pinhas em cada uma das plantações. Para as pinhas amostradas determina o peso de pinhão em cada pinha.

Os dados foram introduzidos no R nos vectores `ppinhaoA` e `ppinhaoB`, que designam o peso de pinhão nas pinhas da plantação A e o peso de pinhão das pinhas da plantação B, respectivamente, sendo alguns resultados apresentados no Anexo.

a) Considere os resultados referentes à plantação A.

   i) Construa um intervalo de confiança a 95% para o peso médio de pinhão por pinha desta plantação. Justifique convenientemente.

   ii) Com base na alínea anterior, o que pode dizer sobre a afirmação do proprietário relativamente a esta plantação? Justifique.

   iii) Indique, justificando, o que acontece à amplitude do intervalo de confiança obtido em i) para o peso médio de pinhão desta plantação no caso de ter recolhido uma amostra de 100 pinhas. Nota: suponha que não há alteração no valor da variância observada.

b) Considere os dados recolhidos nas duas plantações, A e B.

   i) Relativamente às duas amostras de peso de pinhão (`ppinhaoA` e `ppinhaoB`) indique, justificando, se são amostras independentes ou emparelhadas.

   ii) Será o peso médio de pinhão na plantação A superior ao peso médio do pinhão na plantação B? Justifique convenientemente.

ANEXO

```
>  #ppinhaoA <- peso de pinhão nas pinhas da plantação A
>  #ppinhaoB<- peso de pinhão nas pinhas  da plantação B


> length(ppinhaoA)               > length(ppinhaoB)
[1] 49                             [1] 49


> mean(ppinhaoA)                 > mean(ppinhaoB)
[1] 40.7549                        [1] 25.29694


> var(ppinhaoA)                 > var(ppinhaoB)
[1]  249.6795                      [1] 69.3119


> var(ppinhaoA-ppinhaoB)
[1] 286.8475
```