

Exame 2ª Época – Versão A – 20/07/2015

Curso: _____ Número: _____ Nome: _____

Leia, por favor, com atenção:

1. Este enunciado corresponde à primeira parte do exame (**R**). O enunciado da segunda parte (SAS) será entregue após o aluno entregar a resolução da primeira parte.
 2. Cada parte está cotada para 10 valores e tem a nota mínima de 4 valores.
 3. Esta parte deverá ser realizado com acesso a um computador com o programa **R** instalado. Só poderá aceder ao programa **R** e ao sistema de ajuda do mesmo.
 4. Os dados necessários para resolver os problemas já se encontram disponíveis na sua instalação do **R**.
 5. É proibido o uso de qualquer material de apoio (livros, apontamentos, telemóvel), assim como a troca de qualquer informação com os colegas.
 6. Deverá escrever o seu nome, número e curso no cabeçalho desta folha.
 7. As respostas às questões deverão ser dadas, exclusivamente, na folha do enunciado, no espaço reservado para tal. Estas respostas deverão ser código em **R**.
 8. O não cumprimento de alguma das regras conduzirá à anulação do exame.
 9. A duração do exame, considerando ambas as partes, é de **2 horas**.
-
1. O vector **precip** contém a precipitação (chuva) média anual, em polegadas (in), de 70 cidades norte americanas.
 - (a) Encontre as cidades, e respectivos valores de precipitação, com os 5% valores de precipitação média anual mais elevados. [1 val.]

Solução:

```
> subset(precip, precip >= quantile(precip, 0.95))  
Mobile      Miami New Orleans    San Juan  
67.0        59.8          56.8      59.2
```

- (b) Em Portugal, a unidade de medida mais usada para precipitação é o milímetro. Sabendo que 1 in = 25.4 mm, defina uma função que recebe um vector com valores de precipitação em in, e devolve um vector com valores de precipitação em mm. Converta os valores do vector **precip** para mm. [0.5 val.]

Solução:

```
in2cm = function(precip) {  
  precip * 25.4  
}  
in2cm(precip)
```

- (c) Utilize a função `strsplit` para separar os nomes dos elementos do vector `precip` por palavras (espaços existentes nos nomes). Guarde o resultado numa nova variável denominada `idades`. [1 val.]

Solução:

```
idades = strsplit(names(precip), " ")
```

- (d) Liste os nomes das cidades, existentes no vector `precip`, cujos nomes são compostos por mais de duas palavras. **Sugestão:** utilize a lista `idades`, criada na alínea anterior. [1.5 val.]

Solução:

```
> names(precip)[sapply(idades, function(x) length(x) > 2)]  
[1] "Sault Ste. Marie" "Salt Lake City"
```

2. O objecto `Orange` tem 35 linhas e 3 colunas com registos do crescimento de laranjeiras. As variáveis disponibilizadas são:

Tree `factor` ordenado que indica em que árvore foi efectuada a medição.

age Idade da árvore, dada em dias.

circumference Perímetro do tronco da árvore, em mm.

- (a) Calcule o perímetro médio das laranjeiras, para cada idade registada. Não utilize uma estrutura iterativa. [1 val.]

Solução:

```
by(Orange$circumference, Orange$age, mean)  
  
ou  
  
aggregate(Orange, list(Orange$age), mean)
```

- (b) Assuma que a Taxa de Crescimento (t_c) de uma laranjeira é dada pela Equação 1.

[2 val.]

$$t_c = \frac{\text{circumference}}{\text{age}} \quad (1)$$

Sem recorrer a um ciclo (explicitamente), calcule a t_c em cada registo. Qual é a árvore com maior Taxa de Crescimento média?

Solução:

```
orange_tc = transform(Orange, tc = circumference / age)
> which.max(by(orange_tc$tc, orange_tc$Tree, mean))
4
5
```

A árvore com maior Taxa de Crescimento é a 4.

3. O **data.frame** `LifeCycleSavings` contém informações sobre as poupanças das populações, em 50 países, na década de 1960. Considere apenas as colunas `pop15` e `pop75`, relativas às percentagens da população abaixo de 15 e acima de 75 anos, respectivamente.

- (a) O rácio entre `pop75` e `pop15` (t_e) é um possível indicador de que a população de um país está envelhecida (Equação 2). Crie uma nova tabela, contendo apenas os países em que $t_e > \bar{t}_e + \sigma(t_e)$.

[1 val.]

$$t_e = \frac{\text{pop75}}{\text{pop15}} \quad (2)$$

Solução:

```
> te = LifeCycleSavings$pop75 / LifeCycleSavings$pop15
> subset(LifeCycleSavings, te > mean(te) + sd(te))
sr pop15 pop75      dpi ddpi
Austria      12.07 23.32  4.41 1507.99 3.93
Belgium       13.17 23.80  4.43 2108.47 3.82
Denmark       16.85 24.42  3.93 2496.53 3.99
France        12.64 25.06  4.70 2213.82 4.52
Germany       12.55 23.31  3.35 2457.12 3.44
Italy         14.28 24.52  3.48 1390.00 3.54
Luxembourg    10.35 21.80  3.73 2449.39 1.57
Norway        10.25 25.95  3.67 2231.03 3.62
Sweden         6.86 21.44  4.54 3299.49 3.01
Switzerland   14.13 23.49  3.73 2630.96 2.70
United Kingdom 7.81 23.27  4.46 1813.93 2.01
```

- (b) Construa um gráfico que permita identificar o mesmo conjunto de países (não é necessário desenhar o gráfico na folha).

[0.5 val.]

Solução:

```
barplot(te)
abline(h = mean(te) + sd(te), las = 2, space = .5))
```

Gráfico de barras com uma linha horizontal que permite identificar a condição requerida. Os dois últimos argumentos são opcionais.

4. Para resolver os próximos exercícios, considere o seu número mecanográfico (número de aluno).

(a) Crie um conjunto de 10 números aleatórios, \mathbf{x} , a partir da seguinte distribuição:

[1 val.]

$$Norm(\mu = xxxx, \sigma = yyy)$$

em que $xxxx$ são os primeiros quatro dígitos do seu número mecanográfico, e yyy são os restantes três dígitos.

Solução:

```
x = rnorm(n = 10, mean = 1999, sd = 123)
```

(b) A partir do conjunto criado na alínea anterior, crie um novo conjunto \mathbf{y} , com igual número de elementos, através da seguinte linha de código: [0.5 val.]

```
y = sample(x, 10, rep=T)
```

Verifique se as médias de \mathbf{x} e \mathbf{y} são iguais e explique porquê.

Nota: se não resolveu a alínea anterior, considere $\mathbf{x} = \text{seq}(1234, 4321, \text{length}=10)$.

Solução:

```
> y=sample(x, 10, rep=T)
> mean(x)
[1] 2007.595
> mean(y)
[1] 1995.478
```

As médias são diferentes porque, apesar dos dois vectores terem o mesmo número de elementos, \mathbf{y} foi criado a partir de uma amostragem com reposição, sendo, então, mais provável existirem elementos repetidos.

Pergunta	1	2	3	4	Total
Cotação	4	3	1.5	1.5	10
Cotação obtida					