

MESTRADO EM ENGENHARIA DE SISTEMAS

Aula 1 – Introdução ao R

1. A tabela lista o número de casos de SIDA (em milhares) em 25 grandes cidades dos Estados Unidos durante o ano de 1992.

38.3	6.2	3.7	2.6	2.1
14.6	5.6	3.7	2.3	2.0
11.9	5.5	3.4	2.2	2.0
6.6	4.6	3.1	2.2	1.9
6.3	4.5	2.7	2.1	1.8

Construa o histograma para descrever os dados. Calcule a média, a mediana, a amplitude e o desvio padrão. Qual a proporção de cidades com mais de 10 000 casos de SIDA em 1992?

2. A tabela apresenta, para o ano de 2007, a projeção da população das 30 maiores cidades do mundo (<https://www.citypopulation.de/en/world/agglomerations/>).

Cidade	País	Pop	Cidade	País	Pop
Tokyo	Japan	33.40	Canton	China	14.50
Seoul	South Korea	23.20	Dacca	Bangladesh	13.80
Mexico City	Mexico	22.10	Buenos Aires	Argentina	13.50
New York	USA	21.80	Moscow	Russia	13.40
Bombay	India	21.30	Beijing	China	12.70
Delhi	India	21.10	Rio de Janeiro	Brazil	12.20
Sao Paulo	Brazil	20.40	Tehran	Iran	12.10
Los Angeles	USA	17.90	London	Great Britain	12.00
Shanghai	China	17.30	Istanbul	Turkey	11.70
Osaka	Japan	16.60	Lagos	Nigeria	10.00
Cairo	Egypt	15.90	Paris	France	9.95
Calcutta	India	15.50	Chicago	USA	9.75
Manila	Philippines	15.40	Shenzhen	China	9.00
Jakarta	Indonesia	14.90	Bangkok	Thailand	8.55
Karachi	Pakistan	14.80	Wuhan	China	8.55

- A partir da fonte indicada atualize essa informação selecionando as 30 cidades mais densamente povoadas.
- Construa um histograma e um gráfico de barras para estes dados.

3. Na tabela abaixo são apresentados os valores calóricos de 3 tipos de “hamburgers”.

<i>vegetariano</i>	<i>carne</i>	<i>frango</i>
168	145.5	101.1
77	158.3	105
185.1	132.7	85.3
183.7	163.3	117.8
181.4	171.7	99.9
183.1	166.3	121
181.9	172.2	111.3
175.4	157.8	120.4
172.3	177.7	89.5
170.9	170.9	99.5
184.6	131.1	76.8
166.7	161.8	84.5
154.8	162.9	88.2
195.7	182.1	93.6
174.2	134.7	115.1

Produza as estatísticas descritivas que resumam os dados das 3 amostras (ficheiro *hamb2018.xlsx*).

4. A tabela apresenta os pesos à nascença, em gramas, de 55 bebés (ficheiro *peso_bebes.csv*). Usando a função *sample* aleatórios, selecione, sem reposição, uma amostra de 15 bebés.

(Pode consultar os seguintes sítios sobre números aleatórios

http://en.wikipedia.org/wiki/Random_number_table

<http://www.graphpad.com/quickcalcs/randomN1.cfm>)

3355 3380 3530 3445 3890 4100 3130 3050 3110 1120 3075 3210 3580 4230 2840
3540 1320 2580 2255 3010 3300 2610 3180 4180 2510 3585 3140 2805 3735 4310
3330 3650 2860 3600 3870 2795 3000 3960 2220 3920 3790 3640 3020 2540 3340
2600 3380 2620 3050 3860 3580 3170 2690 3825 2955

- Com base na amostra, calcule a média, a mediana, a moda, a amplitude e o desvio padrão.
- Considere a amostra recolhida. A cada valor adicione 100 g. O que acontece ao valor da média e do desvio padrão?
- Considere a amostra recolhida. Multiplique cada valor por 2. O que acontece ao valor da média e do desvio padrão?

5. Os dados representam 40 medidas do comprimento interorbital de pombos, em mm (ficheiro *pombos.csv*).

12.2	12.9	11.8	11.9	11.6	11.1	12.3	12.2	11.8	11.8
10.7	11.5	11.3	11.2	11.6	11.9	13.3	11.2	10.5	11.1
12.1	11.9	10.4	10.7	10.8	11.0	11.9	10.2	10.9	11.6
10.8	11.6	10.4	10.7	12.0	12.4	11.7	11.8	11.3	11.1

- Defina uma nova variável *c_inter_orb* com os valores do ficheiro.
- Calcule as estatísticas descritivas.
- Apresente o histograma e a caixa de bigodes dos dados.

6. O ficheiro *amostra_bebes.csv* apresenta os valores para ganho de peso (*gpeso*, kg) durante a gravidez, a idade gestacional (*igestpa*, semanas), o sexo (1,2) e o peso (*peso*, gramas) do bebé ao nascer.

- Calcule as estatísticas descritivas apropriadas para todas as variáveis.
- Apresente os histogramas para o ganho de peso e peso.
- Apresente o gráfico de dispersão do peso ao nascer em função das semanas de gestação.
- Qual dos valores representa o sexo masculino, 1 ou 2?

7. Os dados sobre a população portuguesa (ficheiro *pop.csv*) nas tabelas abaixo referem-se ao ano de 1992.

Faixa Etária	População	Pop. Masculina
0 - 14	1971658	1008249
15 - 24	1610146	814553
25 - 64	4938558	2374514
> 65	1342308	557409

Faixa Etária	Óbitos	Óbitos Masculinos
0 - 14	1874	1098
15 - 24	1791	1427
25 - 64	21901	14863
> 65	75595	35914

Use os gráficos que entender mais apropriados para caracterizar a população portuguesa (há mais homens ou mulheres, morrem mais homens ou mulheres).

Entrada de dados

`x=c(44,17.5,6); x=scan() 44,17.5,6 ; data.entry(x=c(NA)); x=read.csv(file="nome.csv")`

Comandos R

<code>mean()</code> média aritmética	<code>barplot()</code> gráfico de barras
<code>median()</code> mediana	<code>hist()</code> histograma de frequências absolutas
<code>var()</code> variância	<code>pie()</code> gráfico de torta
<code>sd()</code> desvio padrão	<code>boxplot()</code> caixa de bigodes
<code>fivenum()</code> resumo de 5 números	<code>simple.hist.and.boxplot()</code> hist+caixa
<code>IQR()</code> distância interquartilica	<code>plot()</code> gráfico x,y
<code>summary()</code> resumo de 5 números	<code>simple.scatterplot(x,y)</code> gráfico de dispersão x,y
<code>sample()</code> amostra aleatória	
<code>package UsingR</code>	