

Exercícios Resolvidos da AULA AbERTA: Introdução à Estatística

Carlos Pinto Machado <2200909@estudante.uab.pt>

19 de dezembro de 2023

Conteúdo

1	Introdução	2
1.1	Notas de Implementação	2
1.2	Recursos associados	2
3	Estatística Descritiva com R	3
3.1	Organização de dados no R	3
3.2	Visualização de dados usando o R	4
3.3	Redução de dados: média aritmética e desvio padrão	5
4	Exercícios	7
4.1	7
4.2	8
4.3	10
4.4	11
	Bibliografia	13
	Lista de Tabelas	13
	Lista de Figuras	13

Capítulo 1

Introdução

No decorrer da AULA AbERTA - "Introdução à Estatística: Estatística Descritiva com R"[2], o recurso formativo era um guia homónimo [3], no qual existem exercícios propostos. O presente documento pretende documentar soluções para o mesmo com recurso à Bibliografia recomendada do mesmo documento[1].

1.1 Notas de Implementação

Com recurso ao manual do R[5], existe uma tendência aparente, que é a manipulação de data frames. Parece ser onexo, sob qual o ecossistema floresce. Por tal foi o foco nas soluções demonstradas, recorrendo regularmente aos recursos disponíveis no site.

Para carregamento dos dados e gestão dos dados, foram utilizados ficheiros csv. Esta abordagem parece ser mais razoável, dado que podemos ter ficheiros com milhares de amostras.

1.2 Recursos associados

Na página do repositório[4] estão os recursos, código e datasets, em zip:

<https://github.com/cpmachado/a2edR/releases/latest>.

Capítulo 3

Estatística Descritiva com R

3.1 Organização de dados no R

```
1 xi <- read.csv(file.path("dataset", "ieedr", "irmaos.csv"))$x
2 df <- as.data.frame(table(xi), responseName = "ni")
3 df <- transform(df, fi = prop.table(df$ni))
4 write.csv(
5   df,
6   file.path("tabela", "ieedr", "ex3_1.csv"),
7   quote = FALSE
8 )
```

Listing 3.1: ex3_1.R

i	x_i	n_i	f_i
1	0	11	0.275
2	1	16	0.4
3	2	9	0.225
4	3	2	0.05
5	4	2	0.05

Tabela 3.1: Frequências simples e relativas de irmãos

3.2 Visualização de dados usando o R

Distribuição de pessoas por idas ao supermercado

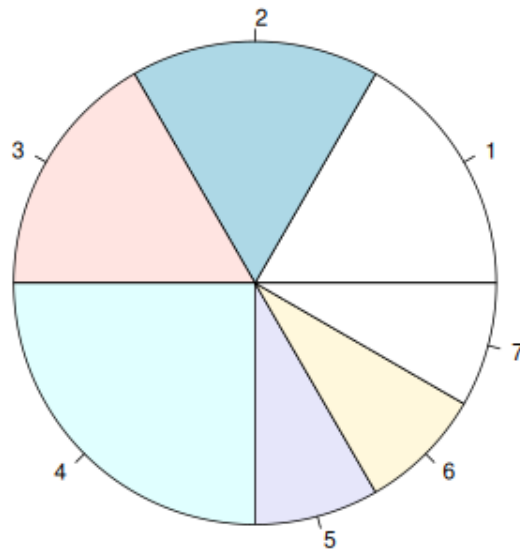


Figura 3.1: Gráfico circular de idas ao supermercado

```
1 png(file.path("imagem", "ieedr", "ex3_2.png"))
2 idas <- read.csv(file.path("dataset", "ieedr", "idas.csv"))
3 pie(
4   table(idas$xi),
5   main = "Distribuição de pessoas por idas ao supermercado"
6 )
```

Listing 3.2: ex3_2.R

3.3 Redução de dados: média aritmética e desvio padrão

$$\bar{x} = 1.2$$

$$s \approx 1.067$$

$$\frac{s}{\bar{x}} \approx 0.889 = 88.9\%$$

A média permite-nos localizar a tendência central em redor dos 1.2 irmãos, e o desvio padrão de aproximadamente 1.067, permite-nos verificar a dispersão associada. Sendo que o quociente do desvio padrão para com a média é relativamente alto, permite-nos concluir que o esbatimento da curva é rápido.

Podemos verificar no gráfico 3.2 esse facto. Aparecem também duas linhas a vermelho, demonstrando o conjunto de amostras no intervalo com centro na média e raio de cumprimento do desvio padrão($\bar{x} \pm s$). Data a natureza discreta da amostra fez-se uns ajustes, sendo o intervalo dado por:

$$[\bar{x} - s] = [1.2 - 1.067] = 1$$

$$[\bar{x} + s] = [1.2 + 1.067] = 2$$

$$I = [[\bar{x} - s], [\bar{x} + s]] = [1, 2]$$

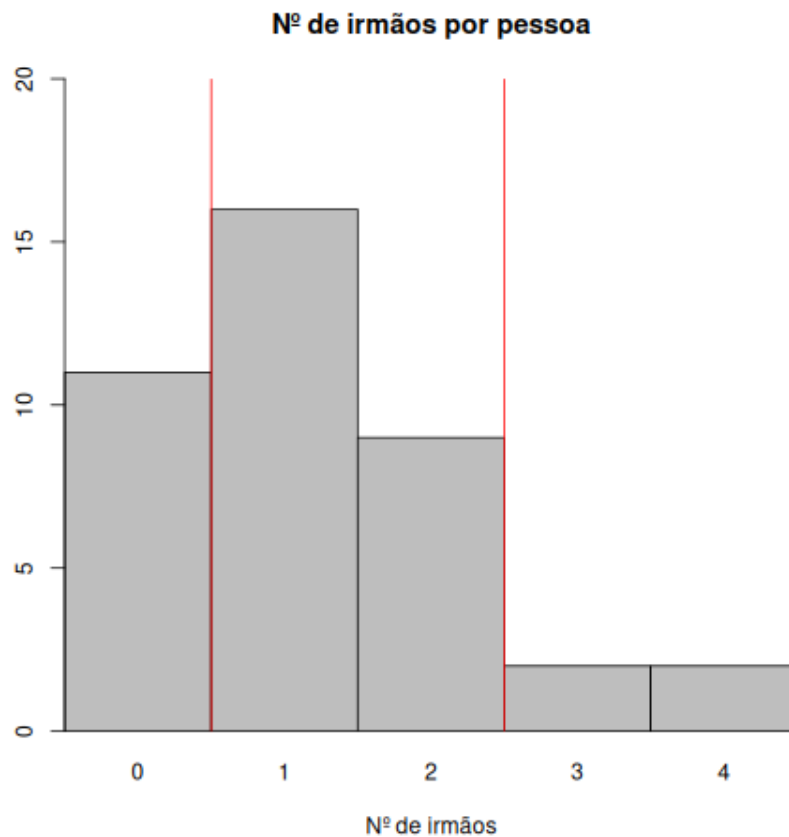


Figura 3.2: Distribuição de irmãos por pessoa

```

1 xi <- read.csv(file.path("dataset", "ieedr", "irmaos.csv"))$x
2 df <- data.frame(media = mean(xi), desvio = sd(xi))
3 df <- transform(
4   df,
5   xmin = ceiling(media - desvio),
6   xmax = floor(media + desvio)
7 )
8 print(df)
9
10 png(file.path("imagem", "ieedr", "ex3_3.png"))
11 barplot(
12   table(xi),
13   xlab = "N° de irmãos",
14   main = "N° de irmãos por pessoa",
15   ylim = c(0, 20),
16   space = 0,
17   yaxs = "i",
18   xaxs = "i"
19 )
20 abline(NULL, v = df$xmin, col = "red")
21 abline(NULL, v = df$xmax + 1, col = "red")

```

Listing 3.3: ex3_3.R

Capítulo 4

Exercícios

Exercícios Propostos

4.1

```
1 t <- read.csv(file.path("dataset", "ieedr", "4_1.csv"))$t
2 df <- data.frame(media = mean(t), desvio = sd(t))
3 print(df)
```

Listing 4.1: ex4_1.R

a) $\bar{x} = 19.41667$

b) $s = 1.378954$

4.2

```

1 df <- read.csv(file.path("dataset", "ieedr", "4_2.csv"), head = TRUE)
2 df <- transform(df, xi = as.numeric(xi))
3
4 # Exercício 4.2 a
5 n <- sum(df$ni)
6 print(paste0("N: ", n))
7
8 # Exercício 4.2 b
9 df <- transform(df, Ni = cumsum(ni), fi = round(prop.table(ni), 3))
10 df <- transform(df, Fi = cumsum(fi))
11 write.csv(
12   df,
13   file.path("tabela", "ieedr", "ex4_2b.csv"),
14   quote = FALSE
15 )
16
17 # Exercício 4.2 c
18 media <- sum(df$xi * df$ni) / n
19 dp <- sqrt(sum((df$xi - media)**2) / (n - 1))
20 print(paste0("média: ", media))
21 print(paste0("desvio padrão: ", dp))
22
23 # Exercício 4.2 d
24 png(file.path("imagem", "ieedr", "ex4_2d.png"))
25 barplot(
26   df$ni,
27   names.arg = as.character(df$xi),
28   ylab = "N° de Colaboradores",
29   xlab = "Salários",
30   main = "Distribuição de salários por colaboradores",
31   ylim = c(0, 14),
32   xaxs = "i",
33   yaxs = "i",
34   space = 0
35 )

```

Listing 4.2: ex4_2.R

a) $N = 60$

b)

i	x_i	n_i	N_i	f_i	F_i
1	1000	10	10	0.167	0.167
2	1100	8	18	0.133	0.3
3	1200	12	30	0.2	0.5
4	1300	8	38	0.133	0.633
5	1400	10	48	0.167	0.8
6	1500	12	60	0.2	1

Tabela 4.1: Tabela de frequências da distribuição de salários

c)

$$\bar{x} = 1260$$

$$s = 54.5552122756449$$

d)

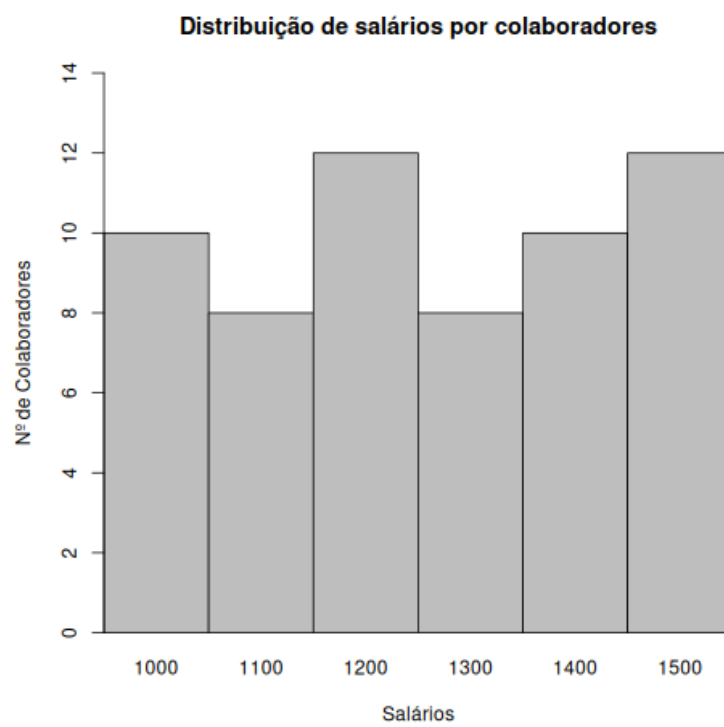


Figura 4.1: Distribuição dos salários por nº de colaboradores

4.3

```
1 p <- read.csv(file.path("dataset", "ieedr", "4_3.csv"))$p
2 df <- data.frame(media = mean(p), desvio = sd(p))
3 print(df)
```

Listing 4.3: ex4_3.R

a) $\bar{x} = 64.26667$

b) $s = 9.676678$

4.4

```

1 df <- read.csv(file.path("dataset", "ieedr", "4_4.csv"))
2
3 # Exercício 4.4 a
4 png(file.path("imagem", "ieedr", "ex4_4a.png"))
5 barplot(
6   df$taxa,
7   names.arg = df$pais,
8   ylab = "Taxa de inflação",
9   xlab = "País",
10  main = "Taxa de inflação por país",
11  ylim = c(0, 2.5),
12  space = 0,
13  xaxs = "i",
14  yaxs = "i"
15 )
16
17 # Exercício 4.4 b
18 png(file.path("imagem", "ieedr", "ex4_4b.png"))
19 pie(
20   df$taxa,
21   labels = df$pais,
22   main = "Taxa de inflação por país"
23 )
24
25 # Exercício 4.4 c
26 print(mean(df$taxa))

```

Listing 4.4: ex4_4.R

a)

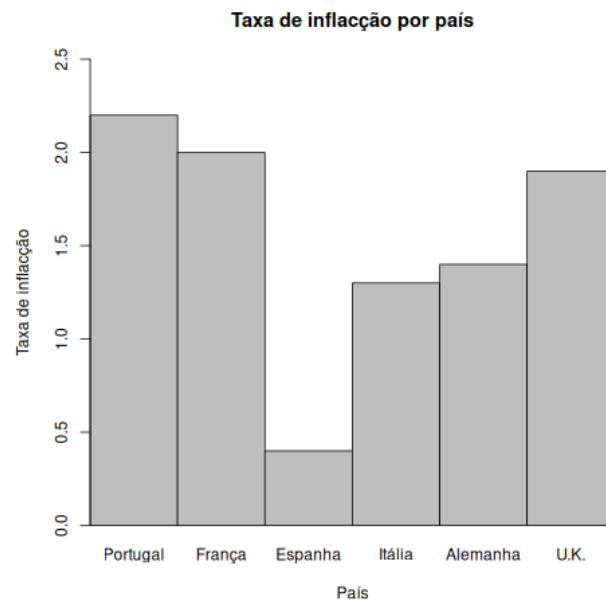


Figura 4.2: Taxa de Inflação por país

b)

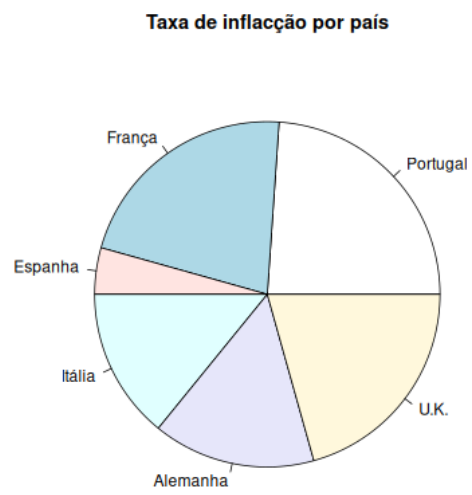


Figura 4.3: Taxa de inflação por país(Gráfico circular)

c) $\bar{x} = 1.533333$

Bibliografia

- [1] Amílcar Oliveira e Teresa A. Oliveira. *Elementos de Estatística Descritiva*. Universidade Aberta, jan. de 2011. URL: <http://hdl.handle.net/10400.2/1986>.
- [2] Teresa A. Oliveira, Amílcar Oliveira e Nuno Sousa. *Introdução à Estatística: Estatística Descritiva em R. AULA AbERTA*. 2017. URL: <https://aulaberta.uab.pt/blocks/catalog/detail.php?id=7>.
- [3] Teresa A. Oliveira, Amílcar Oliveira e Nuno Sousa. *Introdução à Estatística: Estatística Descritiva em R*. Universidade Aberta, 2017. URL: <http://hdl.handle.net/10400.2/12832>.
- [4] Carlos (cpmachado) Pinto Machado. *a2edR. Release section on Github*. URL: <https://github.com/cpmachado/a2edR/releases/latest>.
- [5] R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2022. URL: <https://www.R-project.org/>.

Lista de Tabelas

3.1	Frequências simples e relativas de irmãos	3
4.1	Tabela de frequências da distribuição de salários	8

Lista de Figuras

3.1	Gráfico circular de idas ao supermercado	4
3.2	Distribuição de irmãos por pessoa	5
4.1	Distribuição dos salários por nº de colaboradores	9
4.2	Taxa de Inflação por país	11
4.3	Taxa de inflação por país(Gráfico circular)	12