

Estatística 2020-2021 – *Script* do

```
# -----#
# Script de apoio a uma possível resolução dos exercícios de Introdução ao R
# e Estatística Descritiva
# Todas as linhas que começam com # são comentários
# Esta resolução pressupõe que os ficheiros necessários se encontram na área
# de trabalho do R
#

# -----#
## Exercício 1.1
# -----#
#a)
Ano0 <- c(68, 88, 101, 82, 96, 74, 65, 74, 52, 99)
Ano1 <- c(67, 87, 90, 76, 98, 69, 68, 65, 59, 70)

#b)
length(Ano0);length(Ano1)

#c)
dif <- Ano0-Ano1;dif

#d)
mean(Ano0);median(Ano0);var(Ano0)

# -----#
## Exercício 1.2
# -----#
#a)
nomes <- c("Maria","Joao","Tome","Catarina")
idades <- c(20,18,21,19)
alturas <- c(165,159,180,192)

#b)
cartac <- c(T,F,T,F)

#c)
max_id <- max(idades);max_id
min_id <- min(idades);min_id
```

```

#d)
media_alt <- mean(alturas);media_alt
nomes[alturas>media_alt]

#e)
nomes[cartac]
length(nomes[cartac])

#f)
colegas <- data.frame(nome_c=nomes,idade_c=idades,altura_c=alturas,
                      carta_c=cartac)
colegas                                #agora vemos a tabela que construímos
attach(colegas)
mean(altura_c)

# -----#
## Exercício 1.3
# -----#
#a)
conc <- read.table("concelho.txt",header=T,sep=",",as.is=T)
conc
attach(conc)
#b)
nomec[numfreg==max(numfreg)]

#c)
totc <- length(nomec)
totc

#d)
dist <- table(nomed)
dist

#e)
plot(dist)
barplot(dist)

#f)
mean(dist)
detach(conc)

```

```

# -----#
## Exercício 1.4
# -----#
#a)
v <- sample(1:6,30,rep=T);v

#b) e c)
vbar <- sum(v)/30;vbar
mean(v) #verifique que dá o mesmo
var(v)
s2 <- sum((v-vbar)^2)/29; s2 #0 resultado é o mesmo

#Note-se que obteria um resultado diferente se fizesse
s2exp <- sum(v-vbar)^2/29;s2exp

#d)
table(v)
par(mfrow=c(2,2))#agora vamos dividir o espaço para desenho em 4 regiões
plot(table(v))
w1 <- sample(1:6,300,rep=T)
table(w1)
plot(table(w1))
w2 <- sample(1:6,3000,rep=T)
w2
table(w2)
plot(table(w2))
dev.off()

# -----#
#Exercicio 1.5
# -----#
load("SementesOzono.Rdata")
sementes
ozono
mean(sementes)
mean(ozono)

# -----#
#Exercicio 1.6
# -----#
load("DadosMeteo.Rdata")
precip

```

```

temp
vento
# a)
length(precip[is.na(precip)])
length(temp[is.na(temp)])
length(vento[is.na(vento)])
#
# b)
#indicadores localização
summary(precip); summary(temp); summary(vento)
# indicadores de dispersão
max(precip)-min(precip)
max(temp,na.rm=T)-min(temp,na.rm=T)
max(vento)-min(vento)
quantile(precip, probs =0.75,type = 2)-quantile(precip, probs =0.25,type = 2)
quantile(temp, probs =0.75,na.rm=T,type = 2)-quantile(temp, probs
=0.25,na.rm=T,type = 2)
quantile(vento, probs =0.75,type = 2)-quantile(vento, probs =0.25,type = 2)
var(precip); var(temp, na.rm=T); var(vento)
sd(precip); sd(temp, na.rm=T); sd(vento)

# -----#
#Exercicio 1.8
# -----#
# a)
nematodos <- read.table("nematodos.txt", header=T) # leitura
ni <- table(nematodos$numero) # frequência absoluta
fi <- ni/sum(ni) # frequência relativa
Ni <- cumsum(ni) # frequência absoluta acumulada
Fi <- Ni/sum(ni) # frequência relativa acumulada
cbind(ni,fi,Fi,Ni) # tabela de frequências
plot(table(nematodos)) # grafico de barras
#
# b)
laranjas <- read.table("laranjas.txt",header=T)
table(laranjas$numero) #numero de valores distintos muito elevado
# histograma construído pelo R sem indicação das classes
hist(laranjas$numero)
# histograma com classes definidas pelo utilizador
numSturges <- round(1+log2(length(laranjas$numero))) # regra Sturges
max(laranjas$numero)

```

```

min(laranjas$numero)
(max(laranjas$numero)-min(laranjas$numero))/numSturges # amplitude aproximada
# de cada classe
hist.laranjas <- hist(laranjas$numero,breaks=c(130,160,190,220,250,280,310))
#apresenta o histograma e guarda informação utilizada na sua construção

# visualização simultânea dos dois histogramas
par(mfrow=c(1,2))
hist(laranjas$numero, main="histograma R")
plot(hist.laranjas, main="histograma utilizador")
attach(hist.laranjas)
ni <- counts
fi <- ni/sum(ni)
Ni <- cumsum(ni)
Fi <- Ni/sum(ni)
nclass <- length(ni)
esq <- breaks[1:nclass]
dir <- breaks[2:(nclass+1)]
data.frame(esq,dir,ni,fi,Fi,Ni)
detach(hist.laranjas)
dev.off()

#
# c)
ovelhas <- read.table("ovelhas.txt",header=T)
ovelhas
hist.ovelhas <- hist(ovelhas$peso)
attach(hist.ovelhas)
ni <- counts
fi <- ni/sum(ni)
Fi <- cumsum(fi)
Ni <- cumsum(ni)
esq <- breaks[1:length(ni)]
dir <- breaks[2:(length(ni)+1)]
data.frame(esq,dir,ni,fi,Fi,Ni)
detach(hist.ovelhas)

# -----#
#Exercicio 1.12
# -----#

golfinho <- c(rep(0,17), rep(1,45), rep(2,84), rep(3,52), rep(4,23),

```

```

rep(5,11), rep(6,2), rep(8,1))
#b)
ni <- table(golfinho)
fi <- ni/sum(ni)
Fi <- cumsum(fi)
cbind(ni,fi,Fi)
plot (table(golfinho))
#c)
#medidas de localizacao
mean(golfinho); quantile(golfinho,type=2)
#medidas de dispersao
var(golfinho); sd(golfinho); max(golfinho)-min(golfinho)
quantile(golfinho,probs=0.75,type=2)-quantile(golfinho,probs=0.25,type=2)
#d)
Fi[3]

#
# -----#
#Exercicio 1.15
# -----#
load("Serras.RData")
serras
# a)
# vamos comparar histogramas com diferentes agrupamentos em classes
# histograma construído pelo R sem indicação das classes
par(mfrow=c(2,2))
hist(serras)
# histogramas com classes definidas pelo utilizador
numSturges <- round(1+log2(length(serras))) # regra Sturges número de classes
numSturges
max(serras)
min(serras)
(max(serras)-min(serras))/numSturges # amplitude aproximada de cada classe
hist(serras,breaks=seq(250,2000,by=250),main="amplitude 250, começa em 250")
hist(serras,breaks=seq(290,2280,by=280), main="amplitude 280")
hist(serras,breaks=seq(200,2200,by=250),main="amplitude 250, começa em 200")
dev.off()

# -----#
#Exercicio 1.16
# -----#

```

```

#d)
precip <- c(0,0,0,0,4.7,0.6,17.2,1.4,11.2,1,3.8,0.3,0,0,0.5,7,0,0,3.3,7.6,0.9,0.
3,18.2,4,4.6,22,15.6,0,3.4,0,0)
# visualização simultânea de histograma e caixa de bigodes
par(mfrow=c(1,2))
# histograma construído pelo R sem indicação das classes
hist(precip)
boxplot(precip) # caixa de bigodes
dev.off()
mean(precip)
median(precip)

```

```

# -----#
#Exercicio 1.20
# -----#
#c)
load("FluxoCalor.RData")
ls()
fluxoCalor
attach(fluxoCalor)
total <- c(DIA1,DIA2,DIA3)
max(total)
min(total)
par(mfrow=c(1,2))
hist(total,breaks=seq(-450,0,by=50),main="hist 1")
hist(total,breaks=seq(-425,25,by=50),main="hist 2")
dev.off()
# tabela de frequencias
hist.total <- hist(total,breaks=seq(-425,25,by=50),plot=F)
attach(hist.total)
ni <- counts
fi <- ni/sum(ni)
Fi <- cumsum(fi)
data.frame(mids,ni,fi,Fi)
detach(hist.total)

```

```
# -----#  
#Exercicio 1.34  
# -----#
```

```
load("CalorEspecifico.RData")  
attach(caloresp)  
#a)  
boxplot(y,horizontal=TRUE)  
#b)  
cor(x,y) # coeficiente de correlação entre x e y  
regres <- lm(y~x)  
regres  
summary(regres) # dá informação mais detalhada sobre a  
#recta de regressão de y sobre x  
#c)  
predict(regres,data.frame(x=60))  
detach(caloresp)
```

```
# -----#  
#Exercicio 1.47  
# -----#
```

```
data( ) #este comando permite visualizar um vasto conjunto de dados que  
# o R possui  
anscombe # permite aceder ao conjunto de dados desigando anscombe  
attach(anscombe)  
#a)  
mean(x1);mean(x2);mean(x3);mean(x4)  
var(x1);var(x2);var(x3);var(x4)  
#b)  
cor(x1,y1);cor(x2,y2);cor(x3,y3);cor(x4,y4)  
#c)  
reg1 <- lm(y1~x1)  
reg1  
reg2 <- lm(y2~x2)  
reg2  
reg3 <- lm(y3~x3)  
reg3  
reg4 <- lm(y4~x4)  
reg4  
#d)
```



```
par(mfrow=c(2,2))
plot(x1,y1);abline(reg1)
plot(x2,y2);abline(reg2)
plot(x3,y3);abline(reg3)
plot(x4,y4);abline(reg4)
#e)
plot(x1,residuals(reg1))
plot(x2,residuals(reg2))
plot(x3,residuals(reg3))
plot(x4,residuals(reg4))
detach(anscombe)
#
```