

# Estatística Básica

Wilton de O. Bussab e Pedro A. Morettin

11 jul 2017

## Capítulo 1: Introdução

O R é um software livre de larga utilização em Estatística e Ciência de Dados. O objetivo deste “texto”notebook” é descrever, de forma simples, de que forma alternativa podemos realizar análises presentes no livro Estatística Básica (Bussab e Morettin, 2017) utilizando o R. Por ser de código aberto ele possui um amplo número de implementações que auxiliam em tal tarefa. Sempre que possível, serão descritas diferentes formas de realizar as análises, para que o leitor tenha a opção de escolher aquela que melhor lhe agrade.

**Pacotes (bibliotecas, packages, ou libraries) utilizados neste notebook:**

```
library(akima)
library(boot)
library(bootstrap)
library(colorRamps)
library(MASS)
library(ggExtra)
library(ggplot2)
library(gmodels)
library(knitr)
library(mvtnorm)
library(pander)
library(reshape2)
library(scales)
library(scatterplot3d)
library(stats)
library(stats4)
library(tables)
library(xtable)
```

**Dados e funções programadas ao longo dos capítulos:**

Os dados e funções programadas podem ser acessados em uma imagem do R (arquivo .RData) simplesmente utilizando a seguinte linha de comando (função load) do R.

```
load(url(description = "https://www.ime.usp.br/~pam/dados.RData"))
```

## Capítulo 2: Resumo de Dados

### Tabela 2.1

```
#tab2_1<-read.table("tabela2_1.csv", dec=",", sep=";", h=T)
attach(tab2_1)
names(tab2_1)
```

```
## [1] "N"                "estado_civil"    "grau_instrucao"  "n_filhos"
## [5] "salario"          "idade_anos"      "idade_meses"     "reg_procedencia"
```

```
summary(tab2_1$salario)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  4.000   7.553   10.165   11.122   14.060   23.300
```

## Exemplo 2.2

Tabela 2.2

```
ni<-table(tab2_1$grau_instrucao) # Calcula a tabela de frequências absolutas e armazena o resultado em
fi<-prop.table(ni) # Tabela de frequências relativas (f_i)
p-fi<-100*prop.table(ni) # Porcentagem (100 f_i)
```

```
# Adiciona linhas de total
```

```
ni<-c(ni,sum(ni))
```

```
fi<-c(fi,sum(fi))
```

```
p-fi<-c(p-fi,sum(p-fi))
```

```
names(ni)[4]<-"Total"
```

```
tab2_2<-cbind(ni,fi=round(fi,digits=2),p-fi=round(p-fi,digits=2))
```

```
tab2_2
```

```
##              ni  fi  p-fi
## ensino fundamental 12 0.33 33.33
## ensino médio       18 0.50 50.00
## superior           6 0.17 16.67
## Total              36 1.00 100.00
```

Tabela 2.3:

```
#quebras de linha apenas ilustrativas para facilitar a leitura
```

```
tab2_3<-as.data.frame(
  t(rbind(
    ni=c(650,1020,330,2000),
    p-fi=c(32.5,51,16.5,1)
  ))
  ,row.names =c("Fundamental", "Médio", "Superior", "Total")
)
```

```
tab2_3
```

```
##              ni p-fi
## Fundamental  650 32.5
## Médio        1020 51.0
## Superior      330 16.5
## Total        2000 1.0
```

## Exemplo 2.3

Tabela 2.4

```
ni<-table(cut(tab2_1$salario, breaks = seq(4,24,by=4),right=FALSE)) # Frequencias por categorias
tab2_4 <- rbind(ni, p-fi = 100*prop.table(ni)) # Frequencias relativas em %
#quebras de linha apenas ilustrativas para facilitar a leitura
tab2_4 <- as.data.frame(
  t(cbind(
    tab2_4,
    c(sum(tab2_4[1,]),sum(tab2_4[2,])
  )),row.names =c(colnames(tab2_4),"Total")) #Construcao da tabela
tab2_4<-transform(tab2_4,p-fi=round(p-fi,digits=2))
tab2_4
```

```
##      ni  p-fi
## [4,8)  10 27.78
## [8,12)  12 33.33
## [12,16)  8 22.22
## [16,20)  5 13.89
## [20,24)  1  2.78
## Total   36 100.00
```

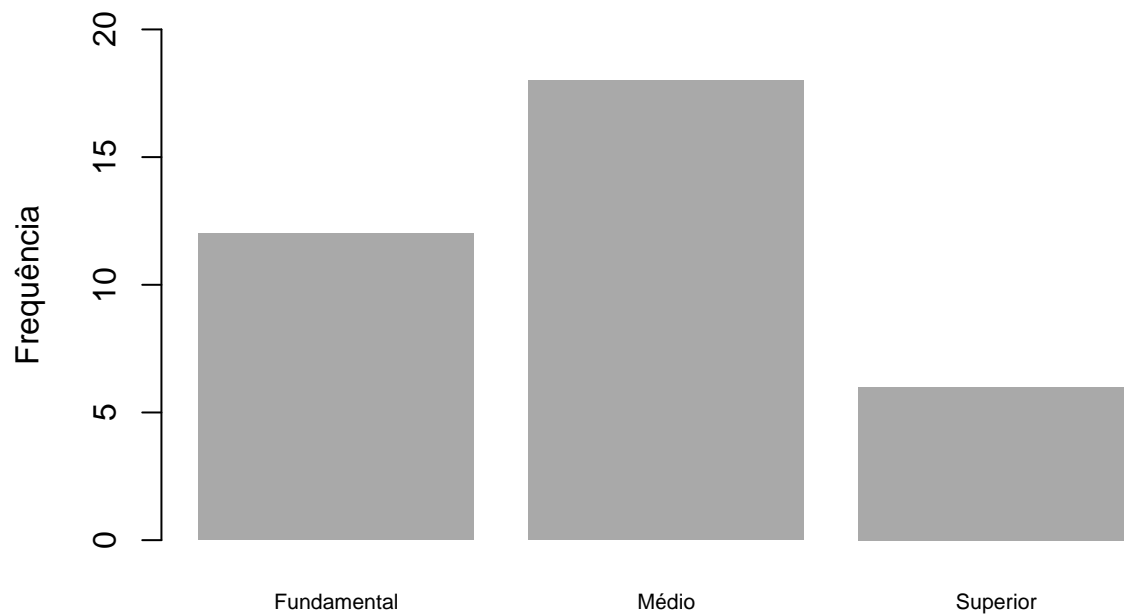
## 2.3 Gráficos

Exemplo 2.4

Figura 2.2

```
#quebras de linha apenas ilustrativas para facilitar a leitura
barplot(
  table(tab2_1$grau_instrucao),
  ylab="Frequência",
  cex.names=0.7,
  names.arg = c("Fundamental","Médio", "Superior"),
  col="darkgrey",
  border=NA,
  main="Figura 2.2: Gráfico em barras para a variável Y: grau de instrução.",
  axes=TRUE,
  ylim=c(0,20)
)
```

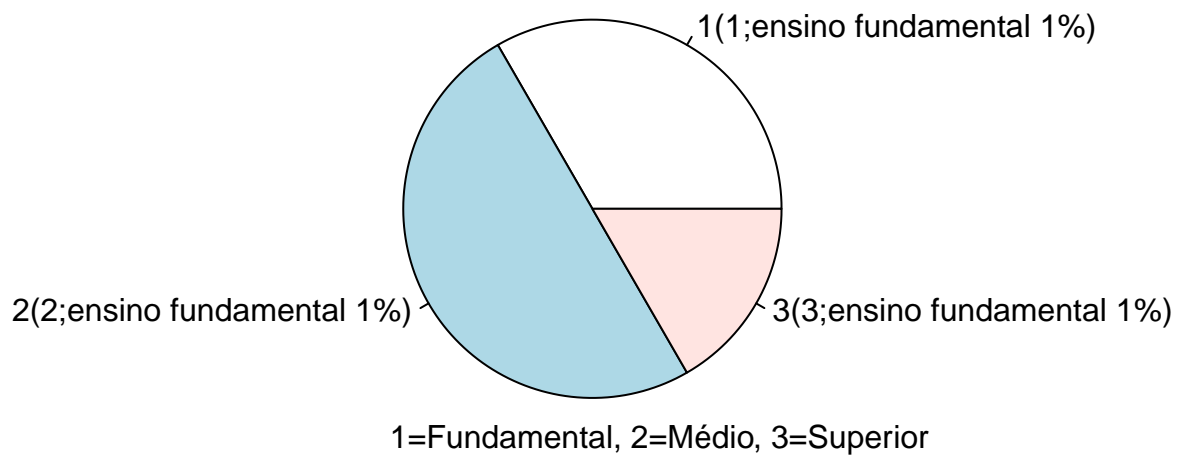
**Figura 2.2: Gráfico em barras para a variável Y: grau de instrução.**



### Figura 2.3

```
labs<-paste(1:3,"(",tab2_1[1:3,1],";",paste(tab2_1[1:3,3],1),"%)",sep="")
pie(table(tab2_1$grau_instrucao),labels=labs)
title("Figura 2.3: Gráfico em setores para a variável Y: grau de instrução")
legend(-1.1,-0.8,legend=c("1=Fundamental, 2=Médio, 3=Superior"),border=NA,box.col=NA)
```

**Figura 2.3: Gráfico em setores para a variável Y: grau de instrução**

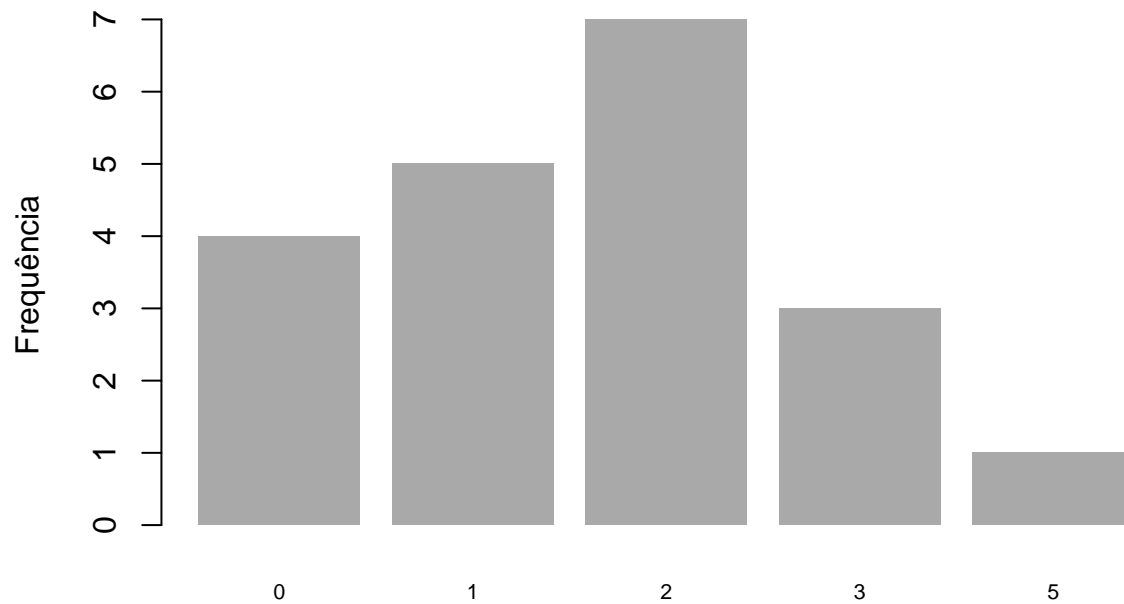


#### Exemplo 2.5

#### Figura 2.4

```
#quebras de linha apenas ilustrativas para facilitar a leitura
barplot(
  table(tab2_1$n_filhos),
  ylab="Frequência",
  cex.names=0.7,
  col="darkgrey",
  main="Figura 2.4: Gráfico em barras para a variável Z: Número de filhos.",
  border=NA)
```

**Figura 2.4: Gráfico em barras para a variável Z: Número de filhos.**



**Figura 2.5**

```
par(mfrow=c(1,3),pin=c(2,2))
tbs<-as.data.frame(cbind(x=0:5,y=c(1,1,1,1,NA,1)),row.names = as.integer(c(4,5,7,3,NA,1)))
plot(tbs,ylim=c(0,7),pch=19,ylab=NA,bty="n",yaxt="n", col="darkblue",xlab="(a)")
text(x=tbs$x,y=tbs$y,rownames(tbs),pos=3)
stripchart(tab2_1$n_filhos,method = "stack", offset = 1, pch = 19, col="darkblue",ylim=c(0,7),ylab=NA,bty="n")
plot(table(tab2_1$n_filhos),type="p", col="darkblue",pch = 19,bty="n",ylab=NA,xlab="(c)")
```

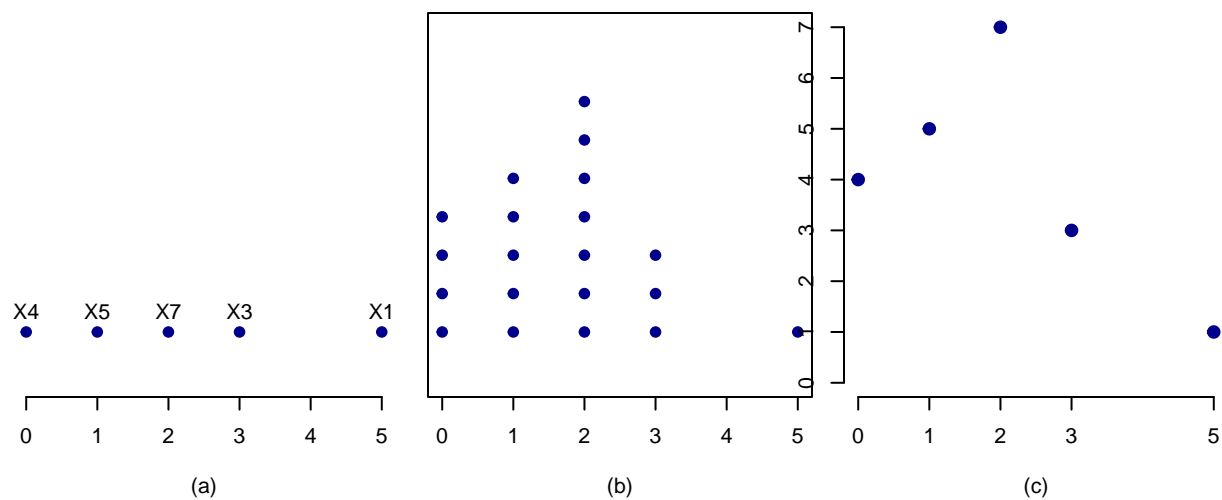


Tabela 2.5

```
ni<-table(tab2_1$n_filhos) # Frequencias absolutas
tab2_5 <- rbind(ni, p-fi = 100*prop.table(ni)) # Frequencias relativas em %
#quebras de linha apenas ilustrativas para facilitar a leitura
tab2_5 <- as.data.frame(
  t(cbind(
    tab2_5,
    c(sum(tab2_5[1,]),sum(tab2_5[2,])
  )),row.names =c(colnames(tab2_5),"Total")) #Construcao da tabela
tab2_5<-transform(tab2_5,p-fi=round(p-fi,digits=2))
tab2_5
```

```
##      ni p-fi
## 0      4  20
## 1      5  25
## 2      7  35
## 3      3  15
## 5      1   5
## Total 20 100
```

## Exemplo 2.6

Figura 2.6

```
#quebras de linha apenas ilustrativas para facilitar a leitura
barplot(
  table(cut(tab2_1$salario, breaks = seq(4,24,by=4),right=FALSE)),
  ylab="Frequência",
  xaxt="n",
  cex.names=0.7,
  col="darkgrey",
  border=NA,
  main="Figura 2.6: Gráfico em barras para a variável S: salários")
axis(1,at=c(.75,1.9,3.1,4.3,5.5),labels=seq(6,22,4),tick=F)
```

Figura 2.6: Gráfico em barras para a variável S: salários

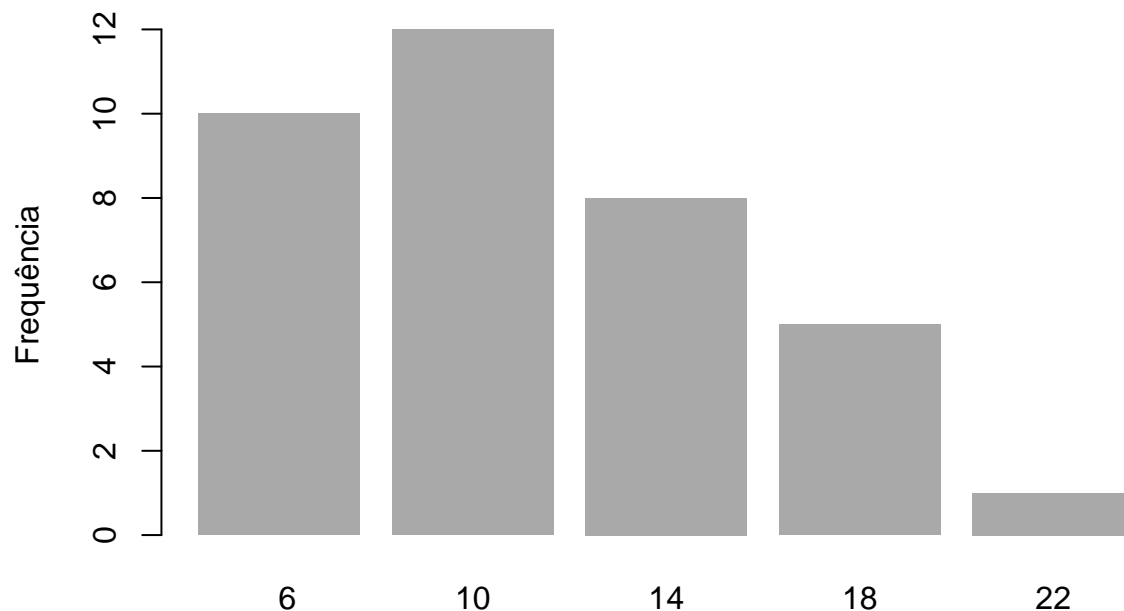


Tabela 2.6

```
ni<-table(cut(tab2_1$salario, breaks = seq(4,24,by=4),right=FALSE)) # Frequências por categorias
#quebras de linha apenas ilustrativas para facilitar a leitura
tab2_6 <- rbind(si=seq(6,22,by=4),ni, p-fi = 100*prop.table(ni)) # Frequências relativas em %
tab2_6 <- as.data.frame(
  t(cbind(tab2_6,
           c(NA,sum(tab2_6[2,]),sum(tab2_6[3,])
           )),row.names =c(colnames(tab2_6),"Total")) #Construção da tabela
tab2_6<-transform(tab2_6,p-fi=round(p-fi,digits=2))
tab2_6
```



```
##          si ni    p_fi
## [4,8)      6 10   27.78
## [8,12)     10 12   33.33
## [12,16)    14  8   22.22
## [16,20)    18  5   13.89
## [20,24)    22  1    2.78
## Total     NA 36 100.00
```

## Exemplo 2.7

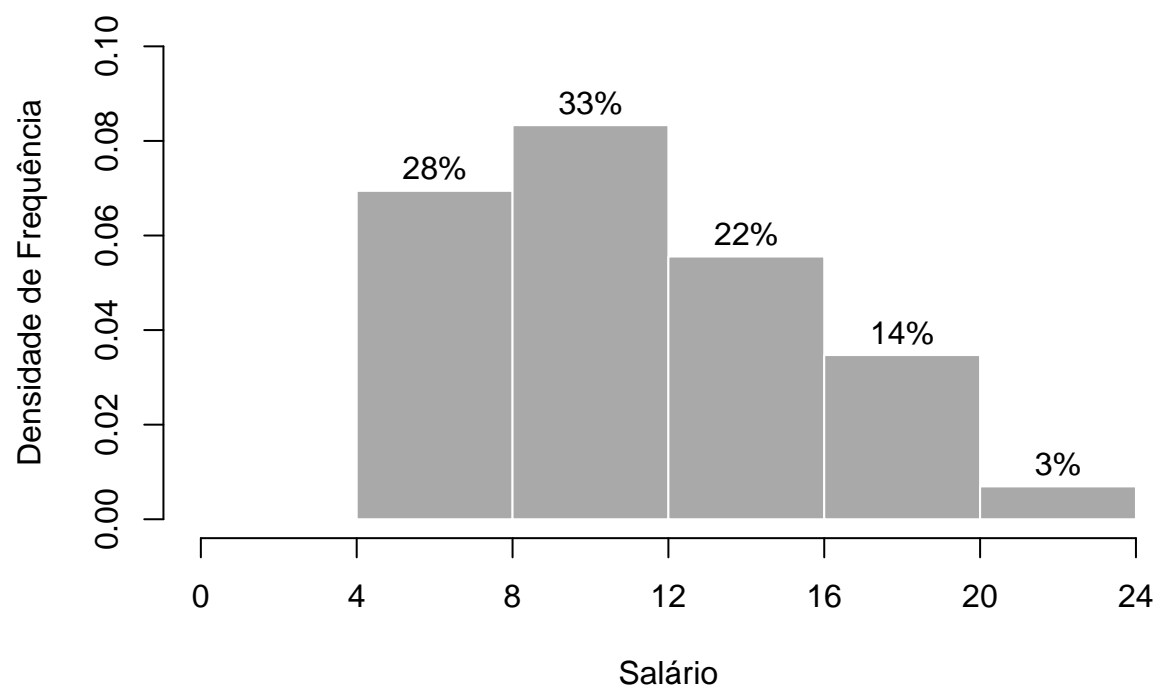
Figura 2.7

```
fig27<-hist(tab2_1$salario, breaks = seq(4,24,by=4),right=FALSE,probability = T,plot=F)

## Warning in hist.default(tab2_1$salario, breaks = seq(4, 24, by = 4), right =
## FALSE, : argumento 'probability' não está sendo utilizado

aux<-with(fig27, 100 * density* diff(breaks)[1])
labs <- paste(round(aux), "%", sep="")
#quebras de linha apenas ilustrativas para facilitar a leitura
plot(fig27,
      freq = FALSE, labels = labs,
      ylab="Densidade de Frequência",
      xlab="Salário",
      col="darkgrey",
      border="white",
      #labels=T,
      main="Figura 2.7: Histograma da variável S: salários",
      xlim=c(0,24), xaxp=c(0,24,6),
      ylim=c(0,.1))
```

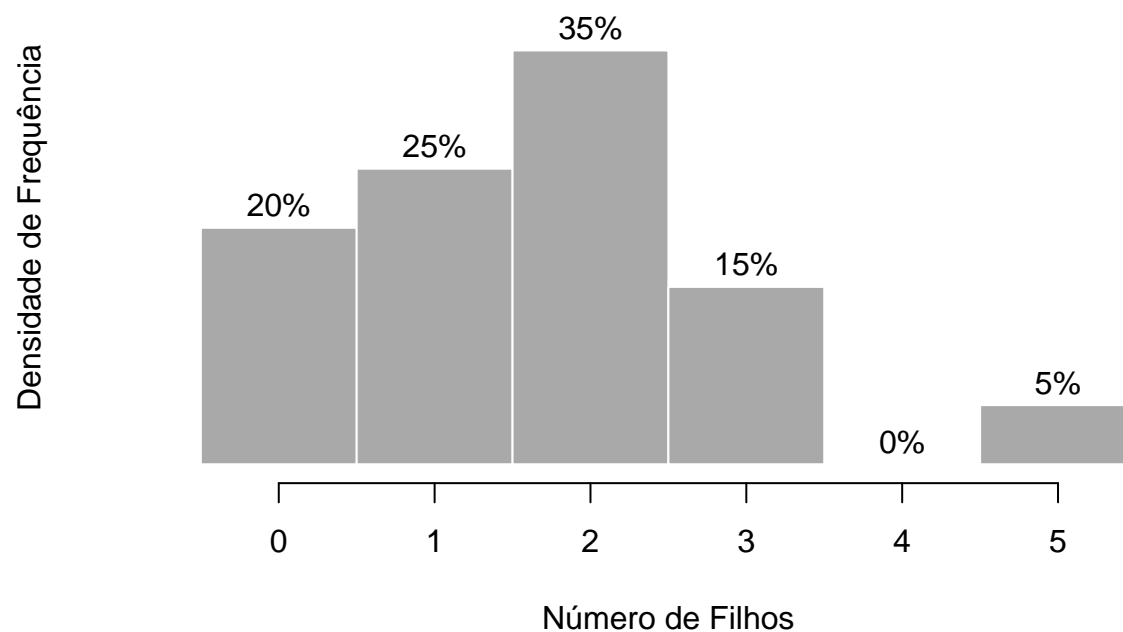
**Figura 2.7: Histograma da variável S: salários**



**Figura 2.8**

```
fig28<-hist(tab2_1$n_filhos, right=F, breaks=seq(-.5,5.5,1),plot=F)
aux<-with(fig28, 100 * density* diff(breaks)[1])
labs <- paste(round(aux), "%", sep="")
#quebras de linha apenas ilustrativas para facilitar a leitura
plot(fig28,
      ylab="Densidade de Frequência",
      xlab="Número de Filhos",
      col="darkgrey",
      border="white",
      bty="n", yaxt="n", ylim=c(0,8),
      main="Figura 2.8: Histograma da variável Z: número de filhos",
      labels=labs)
```

**Figura 2.8: Histograma da variável Z: número de filhos**



## 2.4 Ramo-e-Folhas

### Exemplo 2.8

#### Figura 2.9

```
print("Figura 2.9: Ramo-e-folhas para a Variável S: salários.")
```

```
## [1] "Figura 2.9: Ramo-e-folhas para a Variável S: salários."
```

```
stem(tab2_1$salario,scale=2)
```

```
##
## The decimal point is at the |
##
## 4 | 06
## 5 | 37
## 6 | 379
## 7 | 446
## 8 | 157
## 9 | 01488
## 10 | 58
## 11 | 16
## 12 | 08
## 13 | 269
## 14 | 77
## 15 |
```

```
## 16 | 026
## 17 | 3
## 18 | 8
## 19 | 4
## 20 |
## 21 |
## 22 |
## 23 | 3
```

## Exemplo 2.9

Figura 2.10

```
#quebras de linha apenas ilustrativas para facilitar a leitura
dureza<-c(53 ,70.2,84.3,69.5,77.8,87.5,53.4,82.5,67.3,54.1,
          70.5,71.4,95.4,51.1,74.4,55.7,63.5,85.8,53.5,64.3,
          82.7,78.5,55.7,69.1,72.3,59.5,55.3,73 ,52.4,50.7
        )
stem(as.integer(dureza),scale=.5)
```

```
##
## The decimal point is 1 digit(s) to the right of the |
##
## 5 | 01233345559
## 6 | 34799
## 7 | 00123478
## 8 | 22457
## 9 | 5
```

Figura 2.11

```
print("Figura 2.11: Ramo-e-folhas para dados de dureza, com ramos divididos.")
```

```
## [1] "Figura 2.11: Ramo-e-folhas para dados de dureza, com ramos divididos."
```

```
stem(as.integer(dureza),scale=1)
```

```
##
## The decimal point is 1 digit(s) to the right of the |
##
## 5 | 0123334
## 5 | 5559
## 6 | 34
## 6 | 799
## 7 | 001234
## 7 | 78
## 8 | 224
## 8 | 57
## 9 |
## 9 | 5
```

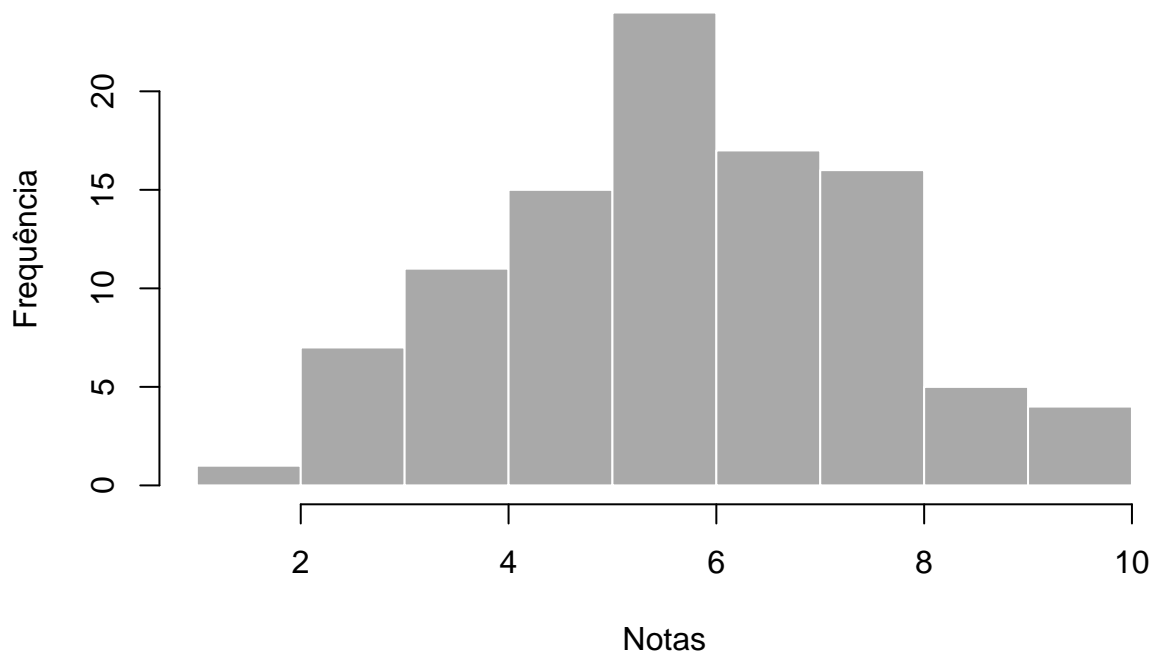
## 2.5 Exemplos Computacionais

### Exemplo 2.10

Figura 2.12

```
#cd_notas<-read.table("cd-notas.csv",h=T,skip=4,sep=";",dec=",")
attach(cd_notas)
hist(cd_notas$nota,col="darkgrey",
     main="Figura 2.12: Histograma para o CD-Notas. R.",
     xlab="Notas",ylab="Frequência",border="white")
```

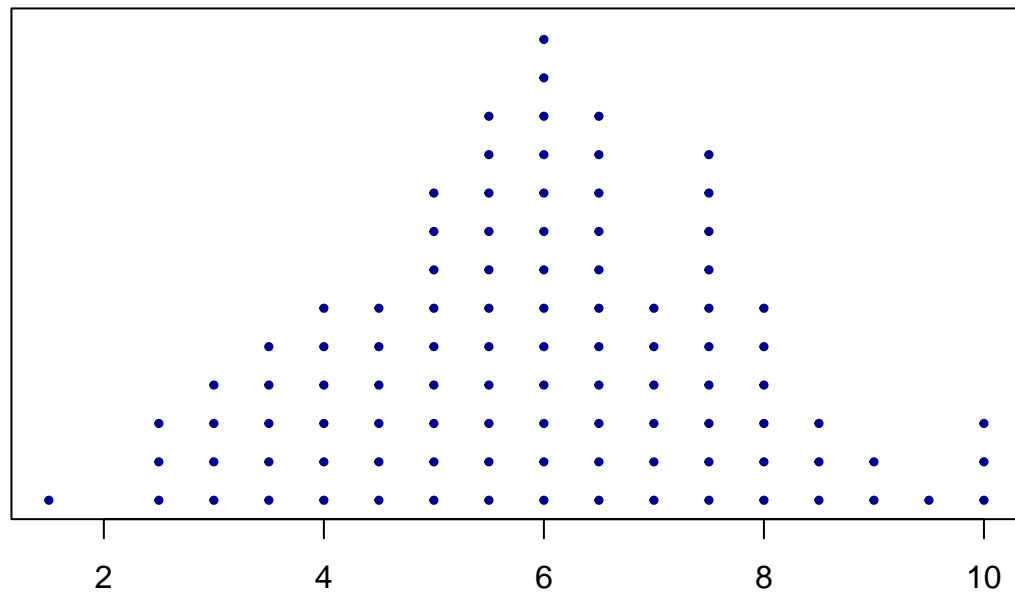
**Figura 2.12: Histograma para o CD-Notas. R.**



### Figura 2.12: Histograma para o CD-Notas. R. ##### Figura 2.13

```
stripchart(cd_notas$nota,method = "stack", offset = 2, at=0,
          main="Figura 2.13: Gráfico de dispersão unidimensional para CD-Notas. R.",
          pch = 19, col="darkblue",ylab=NA,cex=0.5)
```

**Figura 2.13: Gráfico de dispersão unidimensional para CD-Notas. F**



**Figura 2.14**

```
stem(cd_notas$nota)
```

```
##
## The decimal point is at the |
##
## 1 | 5
## 2 | 555
## 3 | 000055555
## 4 | 000000555555
## 5 | 0000000005555555555
## 6 | 00000000000005555555555
## 7 | 0000005555555555
## 8 | 000000555
## 9 | 005
## 10 | 000
```

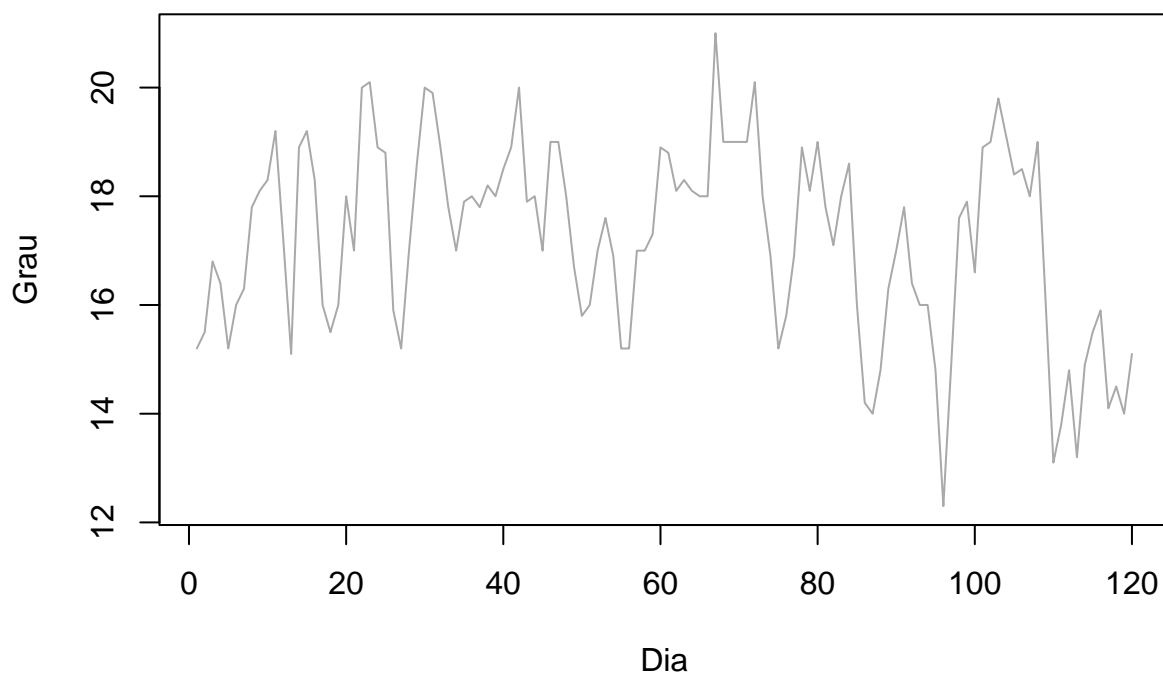
## Exemplo 2.11

**Figura 2.15**

```
#cd_poluicao<-read.table("cd-poluicao.csv",h=T,skip=8,sep=";",dec=",")
attach(cd_poluicao)
#quebras de linha apenas ilustrativas para facilitar a leitura
plot.ts(cd_poluicao$temp,
```

```
main="Figura 2.15: Dados de Temperatura de São Paulo. R.",  
xlab="Dia", ylab="Grau",col="darkgrey")
```

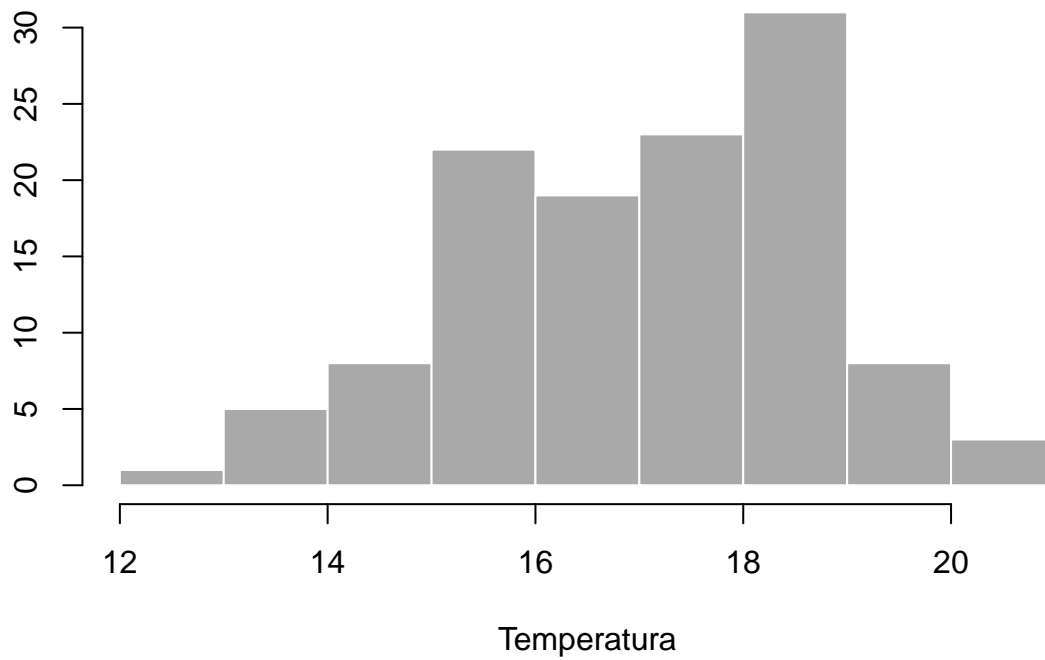
**Figura 2.15: Dados de Temperatura de São Paulo. R.**



**Figura 2.16**

```
#quebras de linha apenas ilustrativas para facilitar a leitura  
hist(cd_poluicao$temp,col="darkgrey",xlab="Temperatura",border="white",  
main="Figura 2.16: Histograma dos dados de Temperatura de São Paulo. R.",  
ylab="")
```

**Figura 2.16: Histograma dos dados de Temperatura de São Paulo. F**

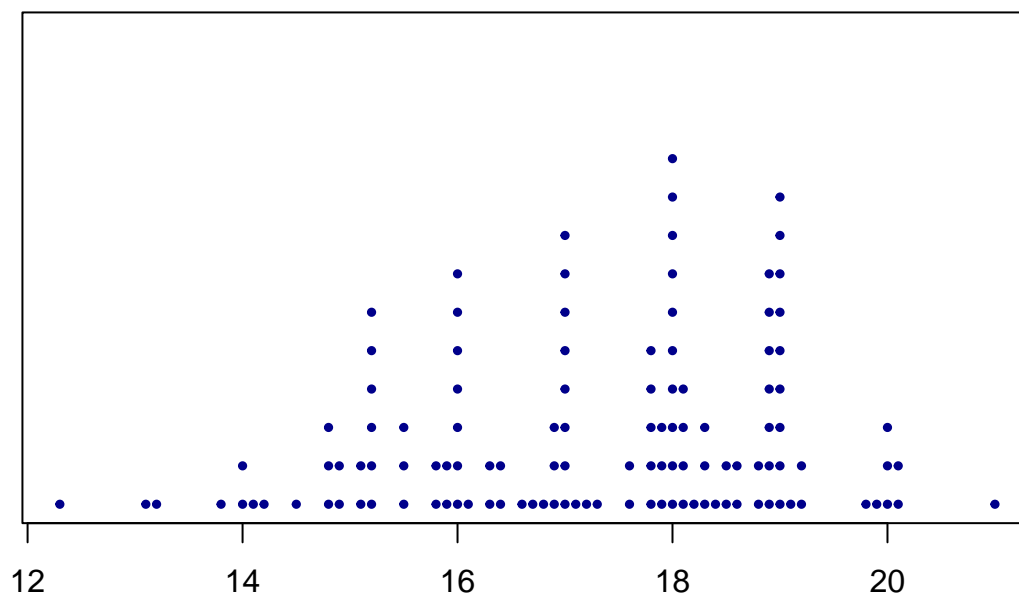


**Figura 2.17**

```
stripchart(cd_poluicao$temp, method = "stack", offset = 2, at=0,  
          main="Figura 2.17: Gráfico de dispersão unidimensional \n para os dados de temperatura de São  
          pch = 19, col="darkblue",ylab=NA,cex=0.5)
```



**Figura 2.17: Gráfico de dispersão unidimensional para os dados de temperatura de São Paulo. R.**



**Figura 2.18**

```
print("Figura 2.18: Ramo-e-Folha para os dados de temperatura de São Paulo. R.")
```

```
## [1] "Figura 2.18: Ramo-e-Folha para os dados de temperatura de São Paulo. R."
```

```
stem(cd_poluicao$temp, scale=.5)
```

```
##
## The decimal point is at the |
##
## 12 | 3
## 13 | 128
## 14 | 0012588899
## 15 | 112222225558899
## 16 | 000000013344678999
## 17 | 000000001236688888999
## 18 | 00000000001111233345566889999999
## 19 | 00000000012289
## 20 | 00011
## 21 | 0
```