

Funções

- > **mean(X)** média
- > **var(X)** variância
- > **sd(X)** desvio-padrão
- > **mad(X)** desvio-médio
- > **max(X)** máximo
- > **min(X)** mínimo
- > **range(X)** menor e maior
- > **median(X)** mediana
- > **sort(X)** ordem crescente
- > **quantile(X)** quantil
- > **table(X)** frequência absoluta
- > **prop.table(table(X))** frequência relativa
- > **attach(X)**
- > **str(X)** nos fornece a quantidade de linhas e colunas, e o tipo das variáveis contidas em X
- > **cor(X,Y)**

4) Deseja-se analisar um conjunto de 10 dados. Cada medida consta de dois valores (Ozônio, Temperatura) que correspondem a concentração de Ozônio (média em partes por bilhão) e temperatura ($^{\circ}\text{F}$) em cada um dos dez locais estudados.

| Medição | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-------------|----|-----|----|----|----|----|----|----|----|----|
| Ozônio | 76 | 118 | 84 | 85 | 96 | 78 | 73 | 91 | 47 | 32 |
| Temperatura | 97 | 94 | 96 | 94 | 91 | 92 | 93 | 93 | 87 | 84 |

a) Calcule a média, mediana, variância e desvio padrão amostrais associados à variável Ozônio e Temperatura.

b) Qual o valor máximo e mínimo de cada uma dessas variáveis?

Gráficos

- PLOT

> plot(X,Y, ...)

type= "p" (pontos)

"l" (linhas)

"b" (pontos e linhas)

"h", (linhas verticais)

main="TÍTULO",

xlab="TÍTULO EIXO x",

ylab=" TÍTULO EIXO y",

- HISTOGRAMA

> **hist(X, ...)**

breaks=n, (n=quantidade de intervalos de classe)

freq= TRUE ou **FALSE**, (FALSE para fr)

main="TÍTULO",

xlab="TÍTULO EIXO x",

ylab=" TÍTULO EIXO y",

col="COR DAS COLUNAS (em inglês)" **colors()**

- DIAGRAMA CAULE E FOLHA

> stem(X, ...)

scale = n , (controla o comprimento do diagrama)

=1 padrão

=2 dobro de linhas do padrão, etc.

• BOX PLOT

> boxplot(X, ...)

xlab="TÍTULO EIXO x",

ylab=" TÍTULO EIXO y",

col="COR DE PREENCHIMENTO DA CAIXA"

names=c("nome caixa 1" ,"nome caixa 2")

Arquivo → Novo script

- Executar linha ou seleção.
- Janelas → Dividir lado a lado.
- Salvar.

1) Os dados abaixo referem-se a dureza de 30 peças de alumínio.

| | | | | | |
|------|------|------|------|------|------|
| 53.0 | 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 | 73.0 | 52.4 | 50.7 |

- a)** Faça uma tabela de distribuição de frequências.
- b)** Faça uma representação gráfica para a distribuição de frequências.
- c)** Calcule média, mediana, desvio padrão e quantis.
- d)** Apresente um histograma dos dados.
- e)** Faça um ramo e folhas.
- f)** Faça um box plot.

> #EXERCÍCIO 1

```
> P = c(53, 53.4, 95.4, 53.5, 72.3,  
+ 70.2, 82.5, 51.1, 64.3, 59.5,  
+ 84.3, 67.3, 74.4, 82.7, 55.3,  
+ 69.5, 54.1, 55.7, 78.5, 73,  
+ 77.8, 70.5, 63.5, 55.7, 52.4,  
+ 87.5, 71.4, 85.8, 69.1, 50.7)
```

> #TABELA DE DISTRIBUIÇÃO DE FREQUÊNCIAS

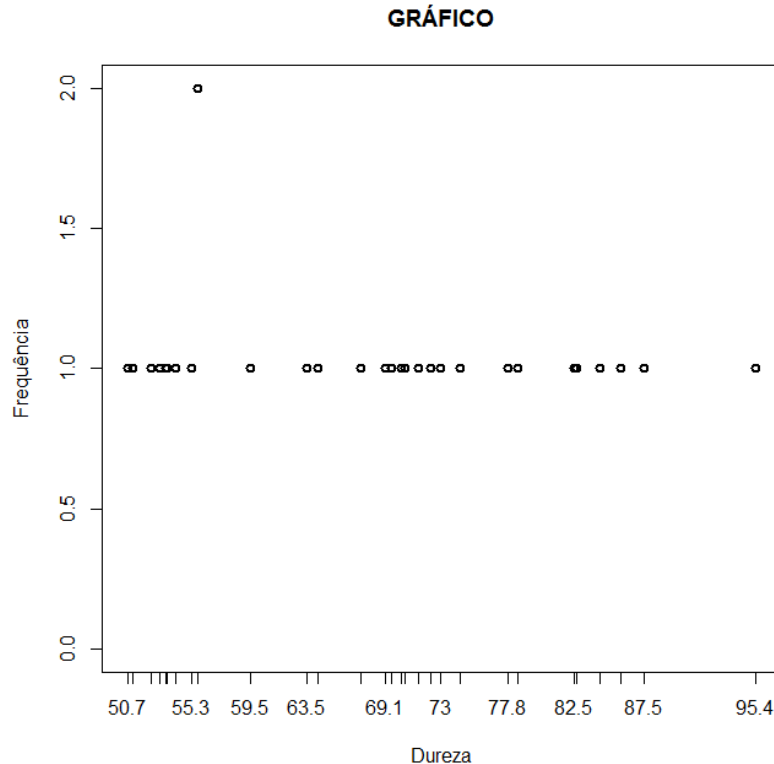
```
> table(P)
```

P

```
50.7 51.1 52.4  53 53.4 53.5 54.1 55.3 55.7 59.5 63.5 64.3 67.3  
69.1 69.5 70.2  
  1  1  1  1  1  1  1  1  2  1  1  1  1  1  1  
70.5 71.4 72.3  73 74.4 77.8 78.5 82.5 82.7 84.3 85.8 87.5 95.4  
  1  1  1  1  1  1  1  1  1  1  1  1  1
```

> #REPRESENTAÇÃO GRÁFICA

```
> plot(table(P), type="p", main="GRÁFICO",  
xlab="Dureza", ylab="Frequência")
```



> #MÉDIA, MEDIANA, DESVIO PADRÃO E QUANTIS

```
> mean(P)
```

```
[1] 67.81333
```

```
> median(P)
```

```
[1] 69.3
```

```
> sd(P)
```

```
[1] 12.72537
```

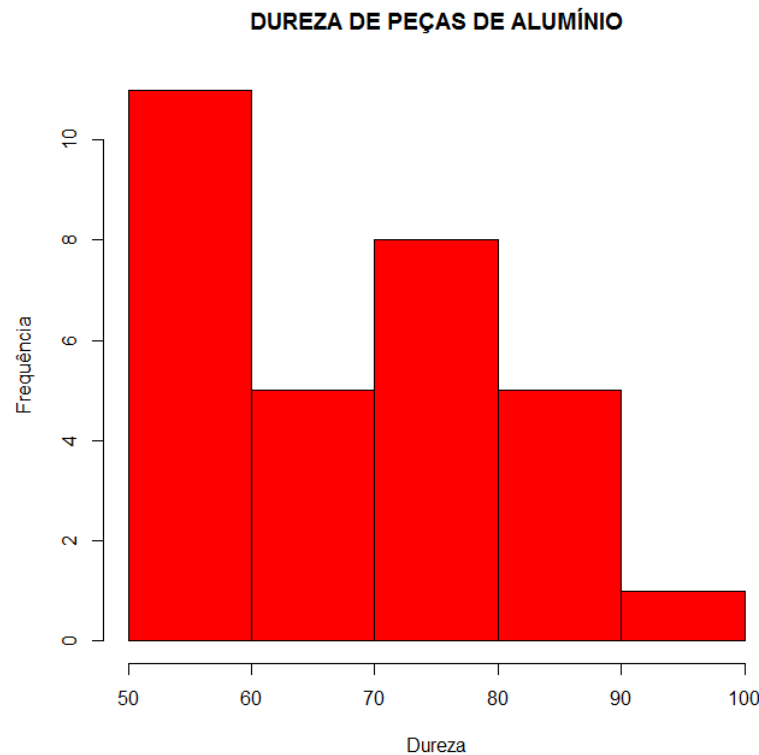
```
> quantile(P)
```

```
0%  25%  50%  75% 100%  
50.70 55.40 69.30 76.95 95.40
```

> #HISTOGRAMA

```
> hist(P, main="DUREZA DE PEÇAS DE  
ALUMÍNIO",
```

```
+ xlab="Dureza", ylab="Frequência", col="red")
```



> #RAMO E FOLHAS

> stem(P)

The decimal point is 1 digit(s) to the right of the |

5 | 1123344

5 | 566

6 | 044

6 | 79

7 | 0011234

7 | 89

8 | 334

8 | 68

9 |

9 | 5

> stem(P, scale=0.5)

The decimal point is 1 digit(s) to the right of the |

5 | 1123344566

6 | 04479

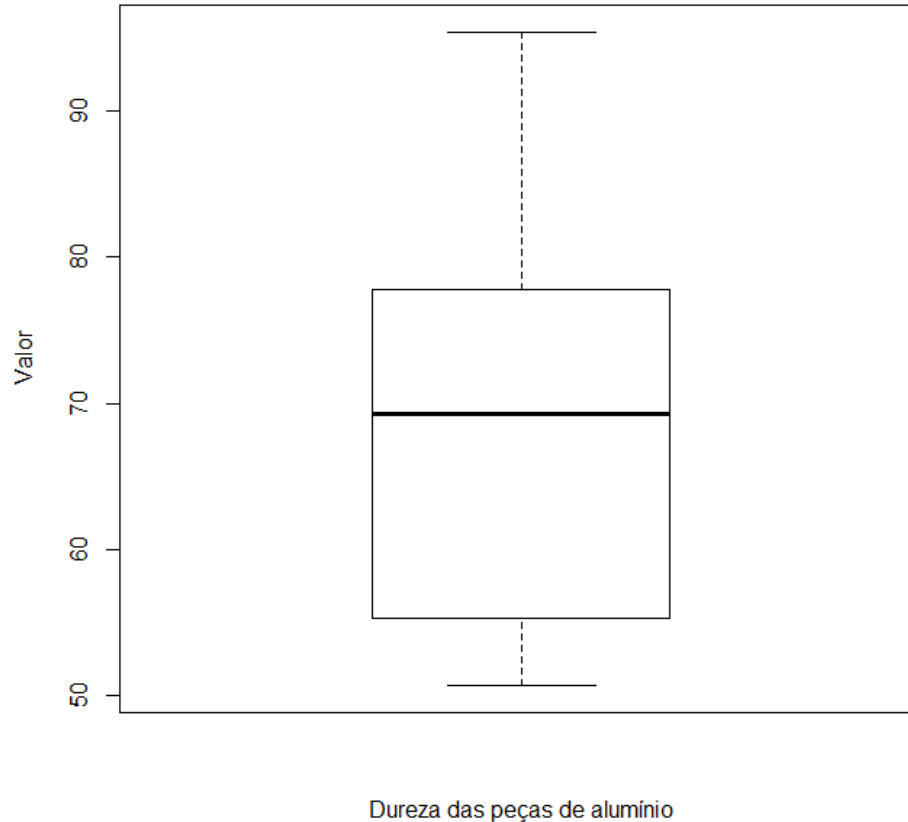
7 | 001123489

8 | 33468

9 | 5

> #BOXPLOT

```
> boxplot(P,xlab="Dureza das peças de  
alumínio",ylab="Valor")
```



2) Considere a altura (em polegadas) de 20 indivíduos.

| | | | | | |
|------------------|-----------|-----------|-----------|-----------|-----------|
| Indivíduo | 1 | 2 | 3 | 4 | 5 |
| Altura | 67.75 | 72.25 | 66.25 | 72.25 | 71.25 |
| Indivíduo | 6 | 7 | 8 | 9 | 10 |
| Altura | 74.75 | 69.75 | 72.5 | 74 | 73.5 |
| Indivíduo | 11 | 12 | 13 | 14 | 15 |
| Altura | 74.5 | 76 | 69.5 | 71.25 | 69.5 |
| Indivíduo | 16 | 17 | 18 | 19 | 20 |
| Altura | 66 | 71 | 71 | 67.75 | 73.5 |

- Faça uma tabela de distribuição de frequências.
- Faça uma representação gráfica para a distribuição de frequências.
- Calcule média, mediana, desvio padrão e quantis.
- Apresente um histograma dos dados.
- Faça um ramo e folhas.
- Faça um box plot.

> #EXERCÍCIO 2

```
> ALTURA = data.frame(  
+ I = c(1:20),  
+ A = c(67.75, 72.25, 66.25, 72.25, 71.25,  
+ 74.75, 69.75, 72.5, 74, 73.5,  
+ 74.5, 76, 69.5, 71.25, 69.5,  
+ 66, 71, 71, 67.75, 73.5))
```

> #TABELA DE DISTRIBUIÇÃO DE FREQUÊNCIAS

```
> table(A)
```

```
#OU
```

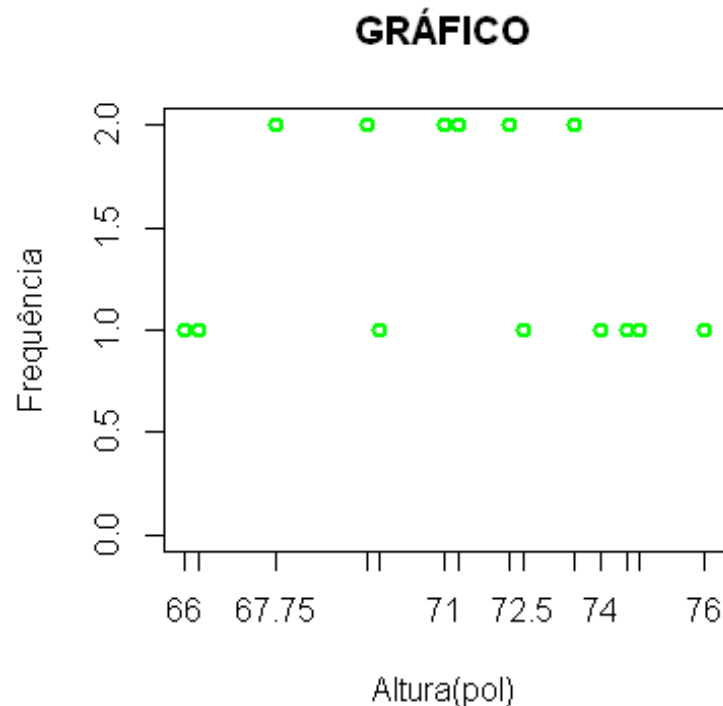
```
> table(ALTURA$A)
```

```
A
```

| | | | | | | | | |
|------|-------|-------|-------|-------|----|-------|-------|------|
| 66 | 66.25 | 67.75 | 69.5 | 69.75 | 71 | 71.25 | 72.25 | 72.5 |
| 1 | 1 | 2 | 2 | 1 | 2 | 2 | 2 | 1 |
| 73.5 | 74 | 74.5 | 74.75 | 76 | | | | |
| 2 | 1 | 1 | 1 | 1 | | | | |

> #REPRESENTAÇÃO GRÁFICA

```
> plot(table(A), type="p", main="GRÁFICO",  
xlab="Altura(pol)", ylab="Frequência",  
col="green")
```



> #MÉDIA, MEDIANA, DESVIO PADRÃO E QUANTIS

> mean(A)

[1] 71.2125

> median(A)

[1] 71.25

> sd(A)

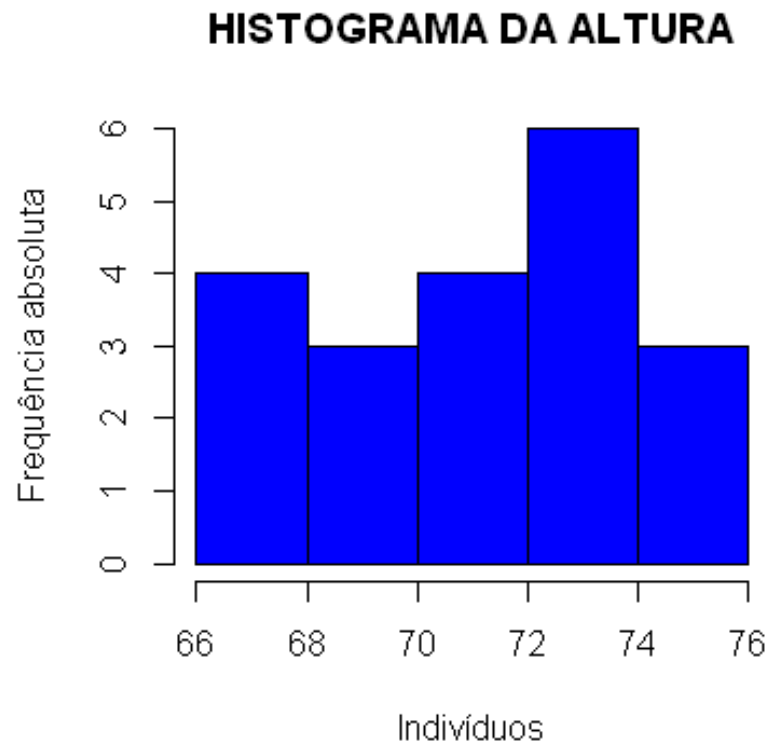
[1] 2.82991

> quantile(A)

| 0% | 25% | 50% | 75% | 100% |
|-------|-------|-------|-------|-------|
| 66.00 | 69.50 | 71.25 | 73.50 | 76.00 |

> #HISTOGRAMA

```
> hist(A, main="HISTOGRAMA DA ALTURA",  
xlab="Indivíduos", ylab="Frequência absoluta",  
col="blue")
```



> #RAMO E FOLHAS

> stem(P, scale=2)

The decimal point is at the |

66 | 0388

68 | 558

70 | 0033

72 | 33555

74 | 058

76 | 0

> stem(P)

The decimal point is 1 digit(s) to
the right of the |

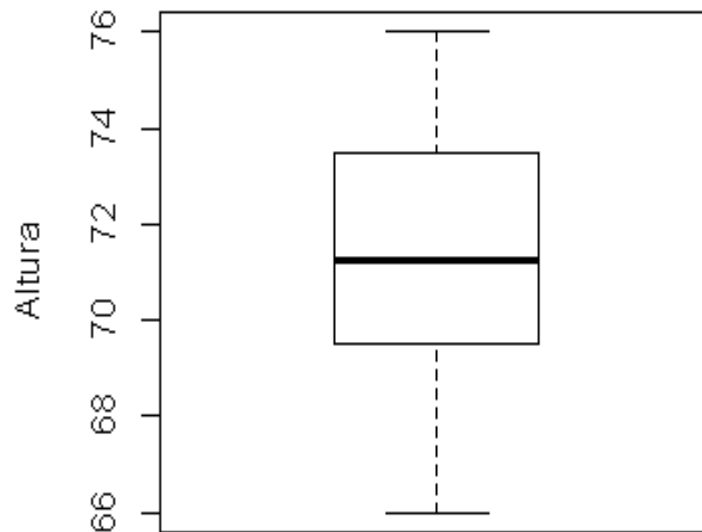
6 | 6688

7 | 0001111223444

7 | 556

> #BOXPLOT

**> boxplot(A,xlab="Altura de 20
individuos",ylab="Altura")**



Altura de 20 individuos

3) As medidas de peso (em libras) e de cintura dos mesmo indivíduos do problema anterior são registradas

| | | | | | |
|-----------|--------|--------|--------|--------|--------|
| Indivíduo | 1 | 2 | 3 | 4 | 5 |
| Peso | 154.25 | 173.25 | 154 | 184.75 | 184.25 |
| Cintura | 94.5 | 98.7 | 99.2 | 101.2 | 101.9 |
| Indivíduo | 6 | 7 | 8 | 9 | 10 |
| Peso | 210.25 | 181 | 176 | 191 | 198.25 |
| Cintura | 107.8 | 100.3 | 97.1 | 99.9 | 104.1 |
| Indivíduo | 11 | 12 | 13 | 14 | 15 |
| Peso | 186.25 | 216 | 180.5 | 205.25 | 187.75 |
| Cintura | 98.2 | 107.7 | 103.9 | 108.6 | 100.1 |
| Indivíduo | 16 | 17 | 18 | 19 | 20 |
| Peso | 162.75 | 195.75 | 209.25 | 183.75 | 211.75 |
| Cintura | 99.2 | 105.2 | 107 | 102.4 | 109 |

b)

i) Calcule a correlação existente entre os seguintes pares de variáveis:

Peso e Altura

Peso e Cintura

Altura e Cintura

ii) Qual das variáveis (Altura ou Cintura) poderia explicar melhor a variável Peso?

iii) Faça um diagrama de dispersão de Cintura e vs Peso.