# Gráficos

## Exemplo

Considere novamente os dados sobre a dureza de 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 |

*Fonte: Hoaglin, Mosteller e Tukey, 1983,* apud *Morettin & Bussab*, Estatística Básica.

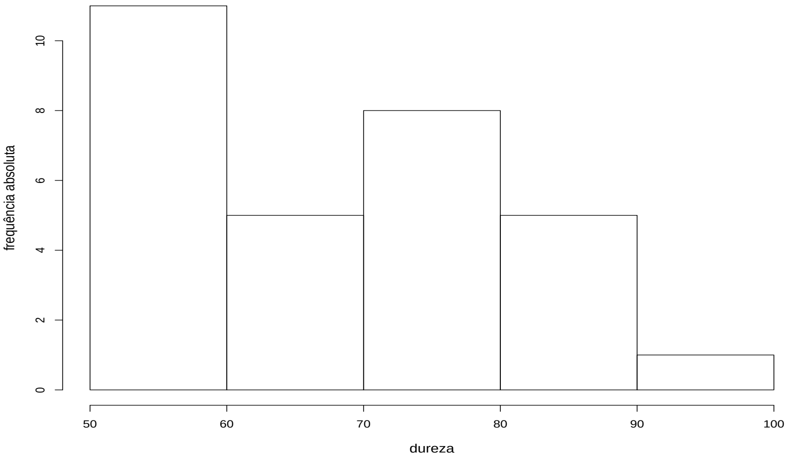
# Gráficos

## Exemplo

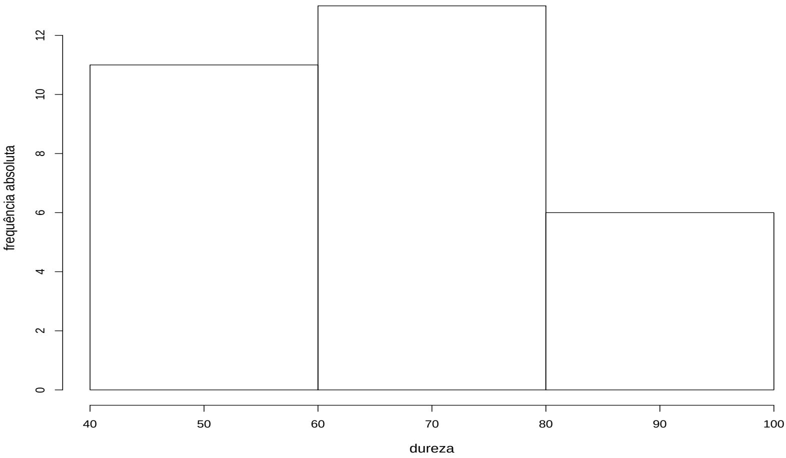
Mostraremos três exemplos de histogramas, além do *boxplot* e do gráfico ramo-e-folhas desses dados. Os histogramas foram gerados com diferentes números de intervalos:

- O primeiro com o padrão do pacote estatístico **R**.
- O segundo com apenas 3 (poucos intervalos).
- O terceiro com 20 (muitos intervalos).

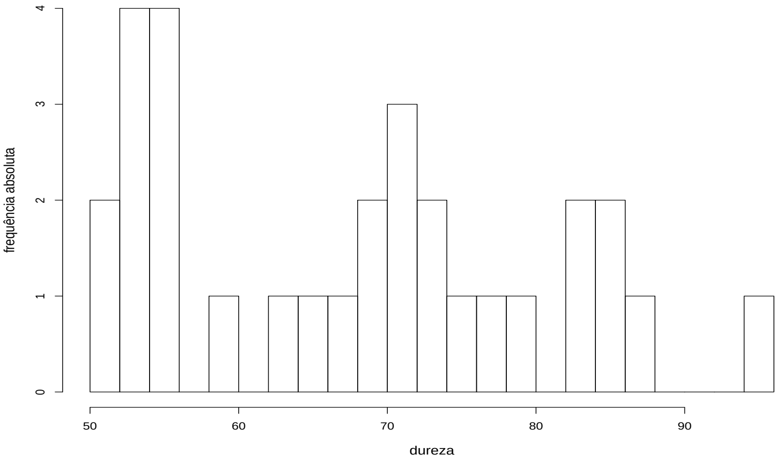
# Histograma



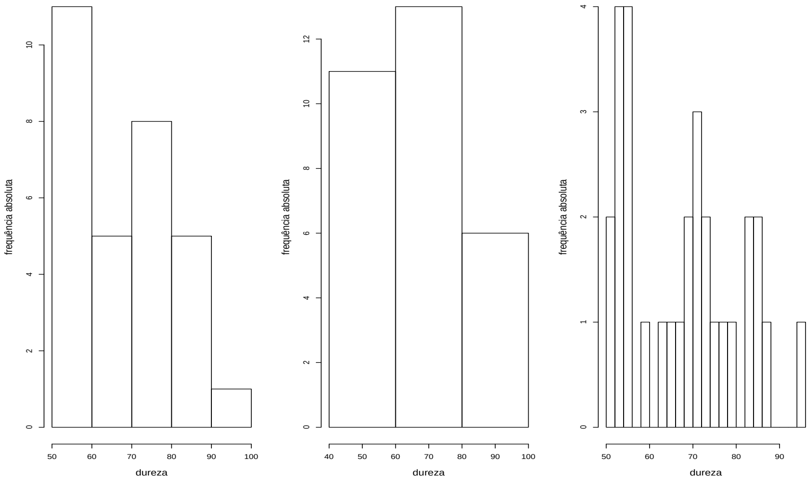
# Histograma (poucas categorias)



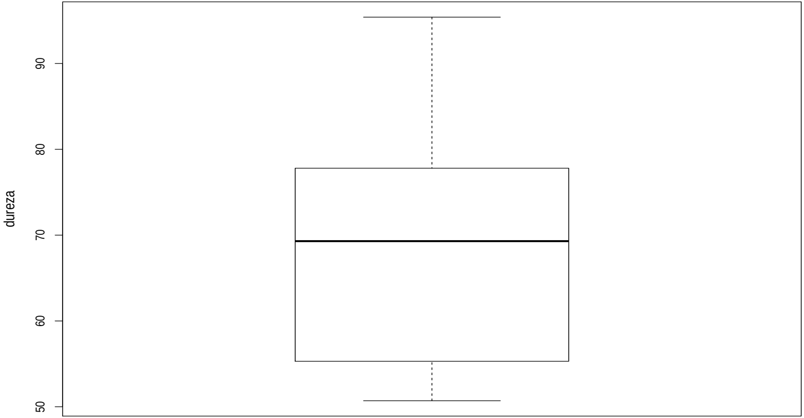
# Histograma (muitas categorias)



# Comparação dos Histogramas



# Box Plot



# Gráfico de Ramo-e-Folhas

O gráfico de Ramo-e-Folhas foi construido para os valores inteiros (truncados) dos dados:

| | |
|---|---|
| 5 | 1 1 2 3 3 4 4 |
| 5 | 5 6 6 |
| 6 | 0 4 4 |
| 6 | 7 9 |
| 7 | 0 0 1 1 2 3 4 |
| 7 | 8 9 |
| 8 | 3 3 4 |
| 8 | 6 8 |
| 9 | |
| 9 | 5 |

# Construção de um Histograma

## Exemplo

A seguinte tabela resume o salário da seção de orçamentos da Companhia MB:

| Salário | Ponto Médio | Frequência | Proporção ($100 f_i$) |
|---|---|---|---|
| [4,00 – 8,00) | 6,00 | 10 | 27,78% |
| [8,00 – 12,00) | 10,00 | 12 | 33,33% |
| [12,00 – 16,00) | 14,00 | 8 | 22,22% |
| [16,00 – 20,00) | 18,00 | 5 | 13,89% |
| [20,00 – 24,00] | 22,00 | 1 | 2,78% |
| Total: | – | 36 | 100% |

*Fonte: Morettin & Bussab*, Estatística Básica 6ª edição, pág 18.

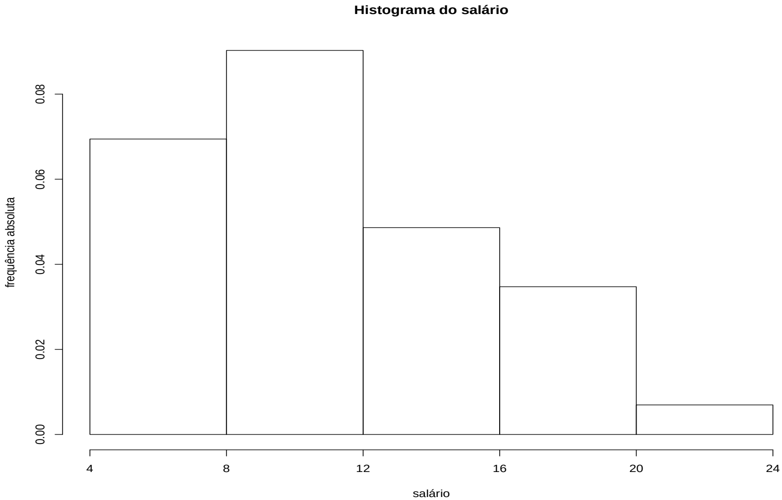
# Construção de um Histograma

O histograma é um gráfico de barras contíguas, onde as bases são proporcionais aos intervalos de classe, e as alturas são correspondem a uma dessas três quantidades:

- Frequência absoluta.
- Frequência relativa.
- Densidade.

Se um certo intervalo tem amplitude $\Delta_i$, então a densidade é dada por $f_i/\Delta_i$, de tal maneira que a área do gráfico seja igual a 1.

# Construção de um Histograma



# Construção de um Box Plot

## Exemplo

Considere a seguinte amostra aleatória de um experimento:

$$0{,}5 \quad 2{,}3 \quad 8{,}0 \quad 9{,}8 \quad 4{,}0 \quad 15{,}3 \quad 6{,}4 \quad 13{,}5 \quad 12{,}0$$

Esses números podem ser ordenados em

$$0{,}5 \quad 2{,}3 \quad 4{,}0 \quad 6{,}4 \quad 8{,}0 \quad 9{,}8 \quad 12{,}0 \quad 13{,}5 \quad 15{,}3$$

*Adaptado de: Morettin & Bussab*, Estatística Básica 6ª edição.

# Construção de um Box Plot

Para construir o Box Plot, devemos determinar algumas estatísticas sobre os dados.

- A *mediana* (ou $Q_2$) que corresponde ao valor central da amostra ordenada, denotada por $x_{(5)}$ neste caso. Seu valor é de 8,0.
- O *primeiro quartil* é o valor mediano (aproximadamente, consoante tivermos um número par ou ímpar de elementos) dos 50% menores valores. Ou seja, o valor mediano de

  $$0{,}5 \quad 2{,}3 \quad 4{,}0 \quad 6{,}4$$

  Como temos um número par de elementos, então o primeiro quartil é a média entre 2,3 e 4,0, ou seja, 3,15.

# Construção de um Box Plot

- O *terceiro quartil* é o valor mediano (aproximadamente, consoante tivermos um número par ou ímpar de elementos) dos 50% maiores valores. Ou seja, o valor mediano de

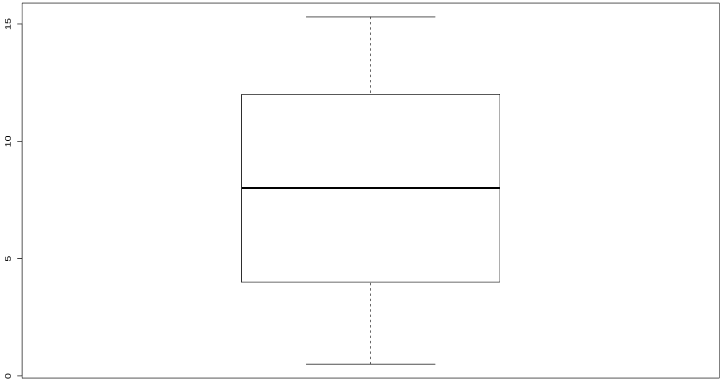
  $$9{,}8 \quad 12{,}0 \quad 13{,}5 \quad 15{,}3$$

  Como temos, novamente, um número par de elementos, então o terceiro quartil é a média entre 12,0 e 13,5, ou seja, 12,75.
- O intervalo interquartilíco $IQ$ é simplesmente $Q_3 - Q_1 = 12{,}75 - 3{,}15 = 9{,}6$.

# Construção de um Box Plot

- A construção do gráfico é imediata: Com os dados no eixo y, o traço horizontal em negrito denota a mediana, a caixa representa a região entre $Q_1$ e $Q_3$, e as linhas pontilhadas denotam o mínimo/máximo dos dados que estiverem na região entre $LIO = Q_1 - 1{,}5IQ$ e $LSO = Q_3 + 1{,}5IQ$.
- Quaisquer valores fora desse intervalo são marcados com um ponto ou asterisco, e chamados *outliers*.
- Neste caso, $LIO = 3,15 - 1,5 * 9,6 = -11,25$ e $LSO = 12,75 - 1,5 * 9,6 = 27,15$.
- Portanto, como $-11,25 < 0{,}5 = \min(x)$ e $\max(x) = 15{,}3 < 27,15$, não temos outliers nos dados.

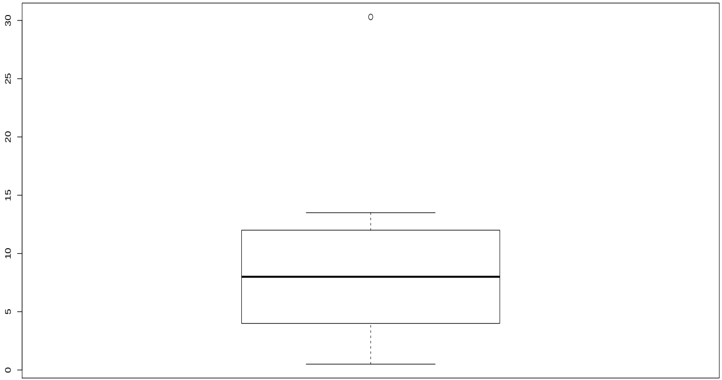
# Construção de um Box Plot



# Construção de um Box Plot

- Podemos perturbar esses valores, para observar o efeito de um outlier nos dados (e como o boxplot consegue detectá-los). Outliers podem ser erros de experimentação, ou inerente ao conjunto de valores (amostra, população) em estudo e devem ser tratados com cautela.
- Suponha que o máximo da amostra tenha sido computado erroneamente, isto é, ao invés de 15,3, computou-se $30,3$. As estatísticas $Q_1$, $Q_2$ e $Q_3$ não se alteram, mas o gráfico resultante exibirá o comportamento destacado dessa observação.

# Construção de um Box Plot



# Distribuições Bivariadas

## Exemplo

Frequentemente os dados serão apresentados em uma tabela, quando lidamos com variáveis discretas, especialmente quando nos interessarem duas ou mais variáveis. Para o caso de duas variáveis $X$ e $Y$, assumindo valores em $1, 2, \ldots, k$ e $1, 2, \ldots, r$, respectivamente, temos que a tabela (de contingência) a seguir é a forma mais adequada de resumir estes dados.

# Distribuições Bivariadas

| | | Y | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | $\ldots$ | r | |
| | 1 | $a_{11}$ | $a_{12}$ | $\cdots$ | $a_{1r}$ | $\sum_{j=1}^{r} a_{2j}$ |
| X | 2 | $a_{21}$ | $a_{22}$ | $\cdots$ | $a_{2r}$ | $\sum_{j=1}^{r} a_{2j}$ |
| | $\vdots$ | $\vdots$ | | $\ddots$ | $\vdots$ | $\vdots$ |
| | k | $a_{k1}$ | $a_{k2}$ | $\cdots$ | $a_{kr}$ | $\sum_{j=1}^{r} a_{kj}$ |
| | | $\sum_{i=1}^{k} a_{i1}$ | $\sum_{i=1}^{k} a_{i2}$ | $\cdots$ | $\sum_{i=1}^{k} a_{ir}$ | $\sum_{i=1}^{k} \sum_{j=1}^{r} a_{ij}$ |

## Exercício

Considere novamente os dados da companhia Milsa (funcionários da empresa).

- Qual a proporção de indivíduos que tem categoria 1 de $Y$ e categoria 2 de $X$?
- Qual a proporção de indivíduos que tem categoria 2 de $Y$, entre o total?
- Qual a proporção de indivíduos que tem categoria 2 de $X$, entre o total?
- Entre os elementos que tem a categoria $r$ de $Y$, que proporção tem a categoria $k$ de $X$?

# Distribuições Bivariadas

## Exemplo

Distribuição conjunta das variáveis **procedência** e **salário**

| | Procedência | | | |
|---|---|---|---|---|
| Salário | Capital | Interior | Outro | Total |
| [4 – 8) | 4 | 3 | 3 | 10 |
| [8 – 12) | 3 | 4 | 6 | 13 |
| [12 – 16) | 1 | 3 | 3 | 7 |
| [16 – 20) | 3 | 1 | 1 | 5 |
| [20 – 24] | 0 | 1 | 0 | 1 |
| Total | 11 | 12 | 13 | 36 |

# Frequências relativas pelo total geral

| | Procedência | | | |
|---|---|---|---|---|
| Salário | Capital | Interior | Outro | Total |
| [4 – 8) | 11,11 | 8,33 | 8,33 | 27,78 |
| [8 – 12) | 8,33 | 11,11 | 16,67 | 36,11 |
| [12 – 16) | 2,78 | 8,33 | 8,33 | 19,44 |
| [16 – 20) | 8,33 | 2,78 | 2,78 | 13,89 |
| [20 – 24) | 0,00 | 2,78 | 0,00 | 2,78 |
| Total | 30,56 | 33,33 | 36,11 | 100,00 |

# Alguns comentários

- Aproximadamente 27,78% dos funcionários ganham entre 4 e 8 salários.
- Aproximadamente 36,11% dos funcionários são de outra região que não Capital/Interior.
- Aproximadamente 11,11% dos funcionários são do interior e ganham entre 8 e 12 salários.
- Aproximadamente 8,33% dos funcionários são da Capital e ganham entre 16 e 20 salários.

# Frequências relativas pelo total por linhas (renda)

| Salário | Procedência | | | Total |
|---|---|---|---|---|
| | Capital | Interior | Outro | |
| [4 – 8) | 40,00 | 30,00 | 30,00 | 100,00 |
| [8 – 12) | 23,08 | 30,77 | 46,15 | 100,00 |
| [12 – 16) | 14,29 | 42,86 | 42,86 | 100,00 |
| [16 – 20) | 60,00 | 20,00 | 20,00 | 100,00 |
| [20 – 24) | 0,00 | 100,00 | 0,00 | 100,00 |
| Total | 30,56 | 33,33 | 36,11 | 100,00 |

# Alguns comentários

- Dos functionários que ganham entre 8 e 12 salário, aproximadamente 23,08% são da capital, 30,77% do interior e 46,15% de outra região.
- Dos que ganham entre 16 e 20, a maioria vem da capital (60%) e aproximadamente a mesma quantidade (20%) vem do interior ou de outra região.
- Dos que ganham entre 20 e 24, 100% vem do interior.
- Muito provavelmente, há uma dependência entre procedência e salário.

# Frequências relativas pelo total por colunas (procedência)

| | Procedência | | | |
|---|---|---|---|---|
| Salário | Capital | Interior | Outro | Total |
| [4 – 8) | 36,36 | 25,00 | 23,08 | 27,78 |
| [8 – 12) | 27,27 | 33,33 | 46,15 | 36,11 |
| [12 – 16) | 9,09 | 25,00 | 23,08 | 19,44 |
| [16 – 20) | 27,27 | 8,33 | 7,69 | 13,89 |
| [20 – 24) | 0,00 | 8,33 | 0,00 | 2,78 |
| Total | 100,00 | 100,00 | 100,00 | 100,00 |

## Alguns comentários

- Dos funcionários provenientes da capital, a maioria (36,36%) ganha entre 4 e 8 salários.
- Dos que vem do interior ou de outra região, a maioria, respectivamente, 33,33% e 46,15%, ganham entre 8 e 12 salários.
- 0,00% dos funcionários que vem da capital ou do interior, tem salário na maior faixa (20 a 24 salários)