

6

Estatística Descritiva

ESQUEMA DO CAPÍTULO

**6.1 IMPORTÂNCIA DO SUMÁRIO E
APRESENTAÇÃO DE DADOS**

6.2 DIAGRAMA DE RAMO E FOLHAS

**6.3 DISTRIBUIÇÕES DE FREQUÊNCIA E
HISTOGRAMAS**

6.4 DIAGRAMA DE CAIXA

6.5 GRÁFICOS SEQUENCIAIS DE TEMPO

6.6 DIAGRAMAS DE DISPERSÃO

6.7 GRÁFICOS DE PROBABILIDADE

Objetivos de Aprendizagem

Após estudo cuidadoso deste capítulo você deverá ser capaz de:

1. Calcular e interpretar a média da amostra, a variância da amostra, o desvio-padrão da amostra, a mediana da amostra e a amplitude da amostra;
2. Explicar os conceitos de média da amostra, variância da amostra, média populacional e variância populacional;
3. Construir e interpretar apresentações visuais de dados, inclusive o diagrama de ramo e folhas, o histograma e o diagrama de caixa (*box-plot*);
4. Explicar como usar diagramas de caixa e outros recursos de apresentação de dados para comparar visualmente duas ou mais amostras de dados;
5. Saber como usar gráficos sequenciais simples para apresentar visualmente características importantes de dados temporais.

6.1 Sumário de Dados

- Média da amostra:

Se as n observações em uma amostra forem denotadas por x_1, x_2, \dots, x_n , então a **média da amostra** será

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

- Exemplo:

Exemplo 6-1

Média da Amostra Vamos considerar as oito observações coletadas nos conectores do protótipo do motor do Capítulo 1. As oito observações são $x_1 = 12,6$, $x_2 = 12,9$, $x_3 = 13,4$, $x_4 = 12,3$, $x_5 = 13,6$, $x_6 = 13,5$, $x_7 = 12,6$ e $x_8 = 13,1$. A média da amostra é

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^8 x_i}{8} = \frac{12,6 + 12,9 + \dots + 13,1}{8} = \frac{104}{8} = 13,0 \text{ libras}$$

Uma interpretação física da média da amostra como uma medida da localização é mostrada no **diagrama de pontos** dos dados da força de remoção. Veja a Figura 6-1. Note que a média da amostra $\bar{x} = 13,0$ pode ser pensada como um “ponto de balanço”. Ou seja, se cada observação representar 1 libra de massa colocada no ponto no eixo x , então o fulcro localizado em \bar{x} equilibrará exatamente esse sistema de pesos.

6.1 Sumário de Dados

- Interpretação física:

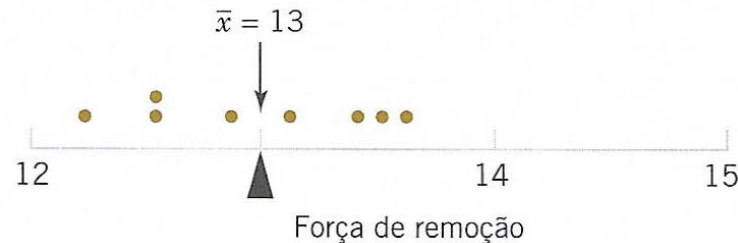


Fig. 6.1 Média da amostra como um ponto de equilíbrio para um sistema de pesos

- Média da população:

$$\mu = \sum_{i=1}^N x_i f(x_i) = \frac{\sum_{i=1}^N x_i}{N}$$

A **média da amostra** é uma **estimativa** razoável da média da população.

6.1 Sumário de Dados

- **Variância da amostra:**

Se x_1, x_2, \dots, x_n for uma amostra de n observações, então a **variância da amostra** será

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad (6-3)$$

O **desvio-padrão da amostra**, s , é a raiz quadrada positiva da variância da amostra.

6.1 Sumário de Dados

- Interpretação da variância:

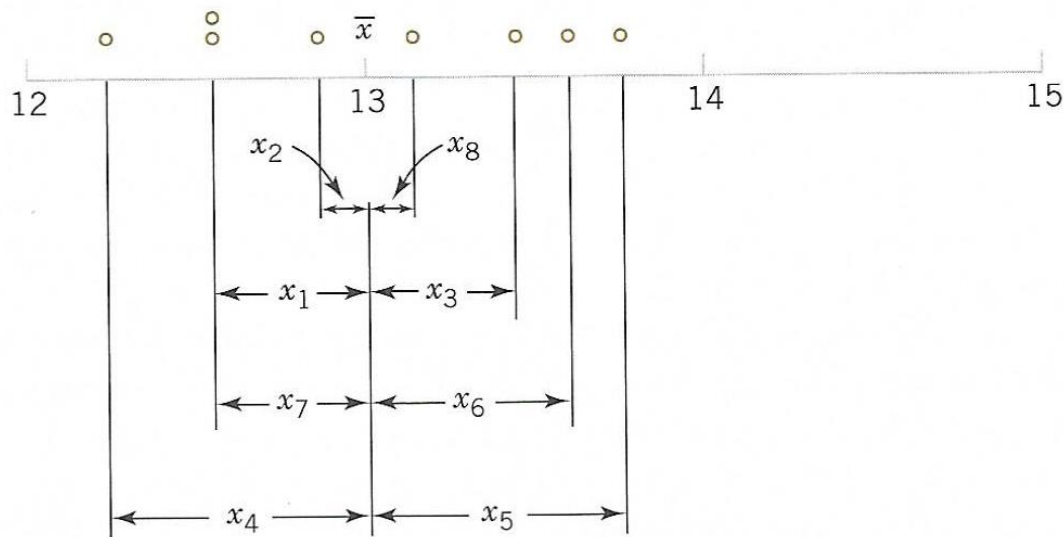


Fig. 6.2 Como a variância mede a variabilidade através dos desvios $x_i - x_{\text{barra}}$

6.1 Sumário de Dados

- Exemplo:

Variância da Amostra A Tabela 6-1 apresenta as grandezas necessárias para calcular a variância e o desvio-padrão da amostra para os dados da força de remoção. Esses dados são plotados na Figura 6-2. O numerador de s^2 é

$$\sum_{i=1}^8 (x_i - \bar{x})^2 = 1,60$$

assim, a variância da amostra é

$$s^2 = \frac{1,60}{8-1} = \frac{1,60}{7} = 0,2286 (\text{libra})^2$$

e o desvio-padrão da amostra é

$$s = \sqrt{0,2286} = 0,48 \text{ libra}$$

6.1 Sumário de Dados

Tab. 6.1 Cálculo dos termos para a variância e desvio-padrão da amostral

i	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	12,6	-0,4	0,16
2	12,9	-0,1	0,01
3	13,4	0,4	0,16
4	12,3	-0,7	0,49
5	13,6	0,6	0,36
6	13,5	0,5	0,25
7	12,6	-0,4	0,16
8	13,1	0,1	0,01
	<u>104,0</u>	<u>0,0</u>	<u>1,60</u>

6.1 Sumário de Dados

- Cálculo **alternativo** de s^2 , que é mais fácil quando feito com calculadoras científicas sem funções estatísticas:

$$s^2 = \frac{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}}{n - 1}$$

6.1 Sumário de Dados

- **Exemplo:** Calcularemos a variância e o desvio-padrão da amostra, usando o método do atalho, Eq. 1.4. A fórmula fornece

$$s^2 = \frac{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}}{n - 1} = \frac{1353,6 - \frac{(104)^2}{8}}{7}$$
$$= \frac{1,60}{7} = 0,2286 \text{ (libras-pé)}^2$$

e

$$s = \sqrt{0,2286} = 0,48 \text{ libras-pé}$$

Esses resultados concordam exatamente com aqueles obtidos previamente.

6.1 Sumário de Dados

- Variância da população:

Quando a população é finita e consiste de N valores, podemos definir a **variância populacional** como

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N} \quad (1.5)$$

A **variância da amostra** é uma estimativa razoável da **variância populacional**.

6.1 Sumário de Dados

- Amplitude da amostra:

Definição

Se as n observações em uma amostra forem denotadas por x_1, x_2, \dots, x_n , então a **amplitude da amostra** será

$$r = \text{máx}(x_i) - \text{mín}(x_i) \quad (1.6)$$

6.2 Diagramas de Ramo e Folhas

- Um **diagrama de ramo e folhas** é uma boa maneira de obter uma apresentação visual informativa de um conjunto de dados x_1, x_2, \dots, x_n , em que cada número x_i consiste em, no mínimo, dois dígitos;
- Passos para construção de um diagrama de ramo e folhas:
 - 1) Dividir cada número em duas partes, (i) um **ramo**, que consiste em um ou mais dígitos iniciais, e (ii) uma **folha**, que consiste nos dígitos restantes;
 - 2) Listar os valores dos ramos em uma coluna vertical;
 - 3) Para cada número, escrever a folha ao lado da respectivo ramo;
 - 4) Escrever as unidades dos ramos e folhas.

6.2 Diagramas de Ramo e Folhas

- **Exemplo:** Para ilustrar a construção de um diagrama de ramo e folhas, considere os dados na Tabela 6.2, sobre a resistência à compressão de uma liga. Como valores dos ramos, selecionaremos os números 7, 8, 9, ..., 24. O diagrama resultante de ramo e folhas é apresentado na Fig. 6.5. A última coluna no diagrama é a frequência do número de folhas associadas a cada ramo. Uma inspeção desse diagrama revela imediatamente que a maioria das resistências à compressão está entre 110 e 200 psi e que um valor central está em algum lugar entre 150 e 160 psi. Além disso, as resistências estão distribuídas aproximadamente de forma simétrica em torno do valor central. O diagrama de ramo e folhas nos capacita a determinar rapidamente algumas características importantes dos dados, que não foram imediatamente óbvias quando da apresentação original na Tabela 6.2.

6.2 Diagramas de Ramo e Folhas

Tab. 6.2 Resistência à compressão de 80 corpos de prova da liga de alumínio-lítio

105	221	183	186	121	181	180	143
97	154	153	174	120	168	167	141
245	228	174	199	181	158	176	110
163	131	154	115	160	208	158	133
207	180	190	193	194	133	156	123
134	178	76	167	184	135	229	146
218	157	101	171	165	172	158	169
199	151	142	163	145	171	148	158
160	175	149	87	160	237	150	135
196	201	200	176	150	170	118	149

6.2 Diagramas de Ramo e Folhas

Fig. 6.4 Diagrama de ramo e folhas para os dados de resistência à compressão na Tab. 6.1

Ramo	Folha	Frequência
7	6	1
8	7	1
9	7	1
10	5 1	2
11	5 8 0	3
12	1 0 3	3
13	4 1 3 5 3 5	6
14	2 9 5 8 3 1 6 9	8
15	4 7 1 3 4 0 8 8 6 8 0 8	12
16	3 0 7 3 0 5 0 8 7 9	10
17	8 5 4 4 1 6 2 1 0 6	10
18	0 3 6 1 4 1 0	7
19	9 6 0 9 3 4	6
20	7 1 0 8	4
21	8	1
22	1 8 9	3
23	7	1
24	5	1

6.2 Diagramas de Ramo e Folhas

- Exemplo:

A Fig. 6.5 ilustra o diagrama de ramo e folhas para 25 observações sobre os rendimentos de uma batelada de um processo químico. Na Fig.6.5 (a), usamos 6, 7, 8 e 9 como os ramos. Isso resulta em muitos poucos ramos e o diagrama de ramo e folhas não provê muita informação sobre os dados. Na Fig.6.5 (b), dividimos cada ramo em duas partes, resultando em uma apresentação mais adequada dos dados. A Fig. 6. 5(c) ilustra um diagrama de ramo e folhas, com cada ramo dividido em cinco partes. Há um número excessivo de ramos nesse gráfico, resultando em um diagrama que não nos diz muito acerca da forma dos dados.

6.2 Diagramas de Ramo e Folhas

Fig. 6.6 Diagrama de ramo e folhas para o exemplo sobre os rendimentos de uma batelada de um processo químico

Ramo	Folha
6	1 3 4 5 5 6
7	0 1 1 3 5 7 8 8 9
8	1 3 4 4 7 8 8
9	2 3 5

(a)

Ramo	Folha
6L	1 3 4
6U	5 5 6
7L	0 1 1 3
7U	5 7 8 8 9
8L	1 3 4 4
8U	7 8 8
9L	2 3
9U	5

(b)

Ramo	Folha
6z	1
6t	3
6f	4 5 5
6s	6
6e	
7z	0 1 1
7t	3
7f	5
7s	7
7e	8 8 9
8z	1
8t	3
8f	4 4
8s	7
8e	8 8
9z	
9t	2 3
9f	5
9s	
9e	

(c)

6.2 Diagramas de Ramo e Folhas

Fig. 6.7 Diagrama de ramo e folhas do Minitab

Diagrama de Caracteres em Ramo e Folhas

Ramo e Folhas da Resistência
N = 80 Unidade da Folha = 1,0

1	7	6
2	8	7
3	9	7
5	10	1 5
8	11	0 5 8
11	12	0 1 3
17	13	1 3 3 4 5 5
25	14	1 2 3 5 6 8 9 9
37	15	0 0 1 3 4 4 6 7 8 8 8 8
(10)	16	0 0 0 3 3 5 7 7 8 9
33	17	0 1 1 2 4 4 5 6 6 8
23	18	0 0 1 1 3 4 6
16	19	0 3 4 6 9 9
10	20	0 1 7 8
6	21	8
5	22	1 8 9
2	23	7
1	24	5

6.2 Diagramas de Ramo e Folhas

- Características dos dados:
 - A **mediana** é uma medida de tendência central que divide os dados em duas partes, metade abaixo da mediana e metade acima. Se o número de observações é par, a mediana está a **meio** caminho entre dois valores centrais;
 - Da Fig. 6.7, o 40º e o 41º valores de resistência são respectivamente 160 e 163,; a mediana é $(160 + 163)/2 = 161.5$; se o número de observações é ímpar, a mediana é o valor **central**;
 - A **amplitude** é uma medida de variabilidade que pode ser facilmente calculada pelo diagrama de ramo e folhas ordenado; é a diferença entre os valores máximo e mínimo;
 - Da Fig. 6.7, a amplitude é $245 - 76 = 169$.

6.2 Diagramas de Ramo e Folhas

- Características dos dados:
 - Quando um conjunto **ordenado** de dados é dividido em quatro partes, os pontos de divisão são denominados **quartis**;
 - O **primeiro quartil** (ou **quartil inferior**), q_1 , é o valor que tem aproximadamente um quarto (25%) das observações abaixo dele e aproximadamente 75% das observações acima;
 - O **segundo quartil** (ou **mediana**), q_2 , tem aproximadamente metade (50%) as observações abaixo do seu valor;
 - O **terceiro quartil** (ou **quartil superior**), q_3 , tem aproximadamente três quartos (75%) das observações abaixo de seu valor;
 - Como no caso da mediana, os quartis podem não ser únicos.

6.2 Diagramas de Ramo e Folhas

- Características dos dados:
 - Os dados da **Tab. 6.2** contêm $n = 80$ observações; o pacote Minitab calcula o primeiro e o terceiro quartis como sendo as $(n + 1)/4$ e $3(n + 1)/4$ observações ordenadas, interpolando quando necessário;
 - Exemplo:
 - $(80 + 1)/4 = 20,25$ e $3(80 + 1)/4 = 60,75$;
 - Assim, o Minitab interpola entre o 20° e 21° observação ordenada, para obter $q_1 = 143,50$ e entre a 60° e 61° observação ordenada para obter $q_3 = 181,00$.

6.2 Diagramas de Ramo e Folhas

- Características dos dados:
 - A **faixa interquartil** é a diferença entre o quartil superior e o quartil inferior e é utilizada algumas vezes como **medida de variabilidade**;
 - Em geral, o $100k^{\circ}$ **percentil** é o valor de modo que aproximadamente $100k\%$ das observações estão nesse ou abaixo desse valor e aproximadamente $100(1-k)\%$ deles estão acima dele.

6.3 Distribuições de Frequência e Histogramas

- Uma **distribuição de frequência** é um sumário mais compacto dos dados, em relação ao diagrama de ramo e folhas;
- Para construir uma distribuição de frequência, precisamos dividir a faixa de dados em intervalos (normalmente de igual amplitude), que são geralmente denominados **intervalos de classes** ou **células**;
- Na prática, o **número** de intervalos é aproximadamente igual à raiz quadrada do número de observações;
 - Exemplo:
Aproximadamente 9 intervalos, para 80 observações da [Tab. 6.2](#).

6.3 Distribuições de Frequência e Histogramas

- Exemplo:

Tab. 6.4 Distribuição de frequências para os dados da [Tab. 6.2](#)

Intervalo de Classe (psi)	Marcação	Frequência	Frequência Relativa	Frequência Relativa Cumulativa
$70 \leq x < 90$		2	0,0250	0,0250
$90 \leq x < 110$		3	0,0375	0,0625
$110 \leq x < 130$		6	0,0750	0,1375
$130 \leq x < 150$		14	0,1750	0,3125
$150 \leq x < 170$		22	0,2750	0,5875
$170 \leq x < 190$		17	0,2125	0,8000
$190 \leq x < 210$		10	0,1250	0,9250
$210 \leq x < 230$		4	0,0500	0,9750
$230 \leq x < 250$		2	0,0250	1,0000

6.3 Distribuições de Frequência e Histogramas

- O **histograma** é uma forma de representação gráfica da distribuição de frequência;
- Passos para **construção** de um histograma de intervalos iguais:
 1. Defina o número de intervalos de classe (aproximadamente a raiz quadrada do número de observações) e calcule e marque os limites das classes na escala horizontal;
 2. Marque na escala vertical as frequências absolutas (ou as relativas);
 3. Sobre cada intervalo de classe, trace um retângulo cuja altura seja proporcional à frequência absoluta (ou frequência relativa) correspondente àquele intervalo.

6.3 Distribuições de Frequência e Histogramas

- Exemplo:

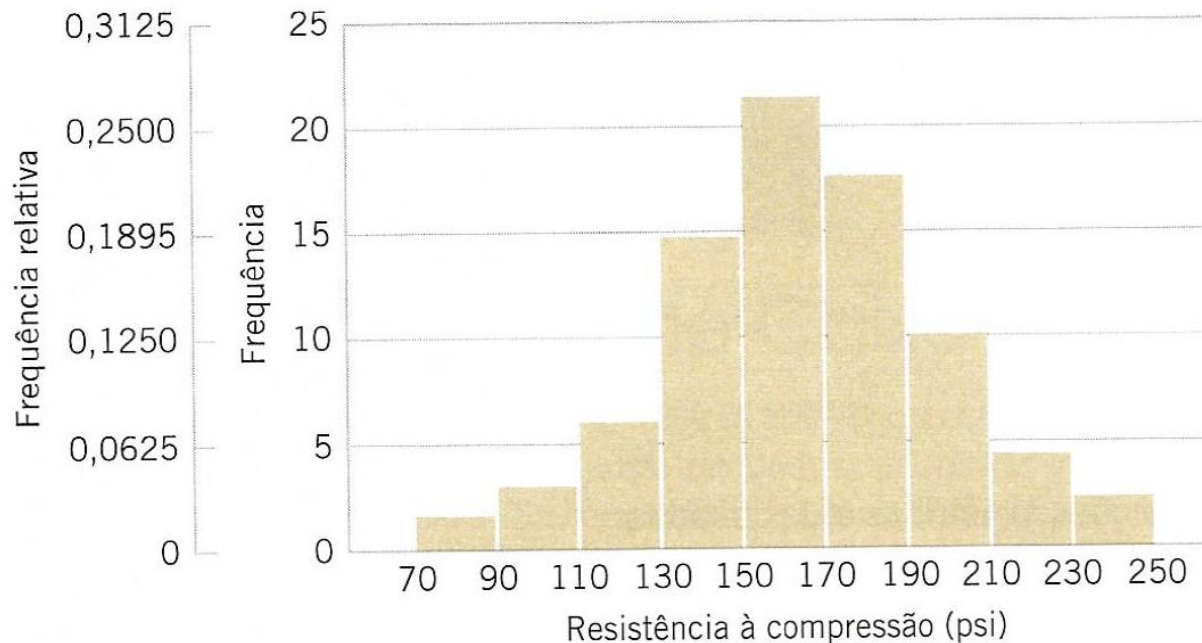


Fig. 6.7 Histograma da resistência à compressão para 80 corpos de prova da liga alumínio-lítio

6.3 Distribuições de Frequência e Histogramas

- Exemplo (cont.):

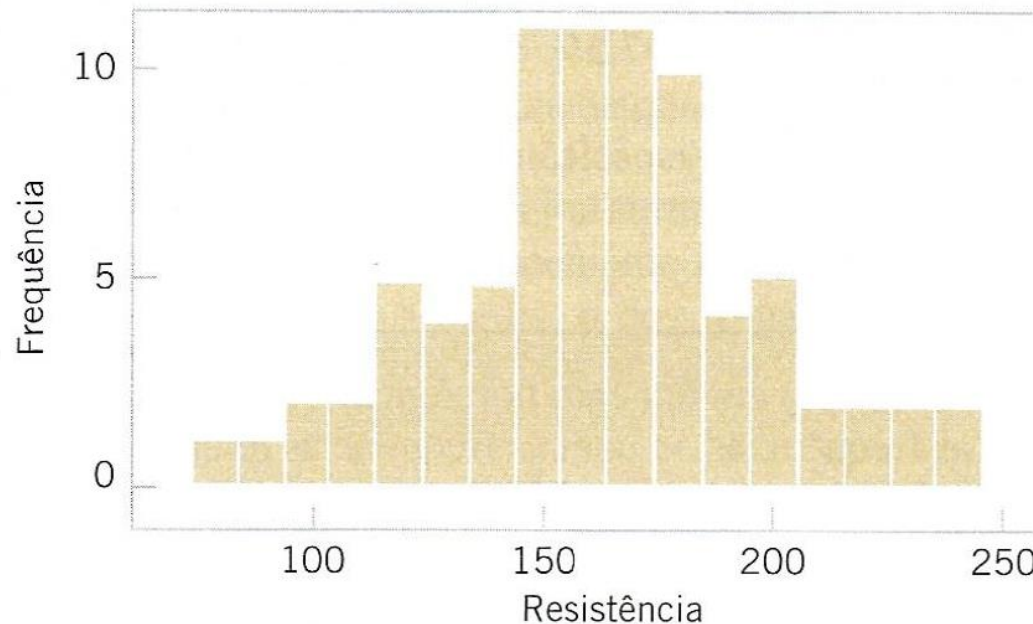


Fig. 6.8 Um histograma dos dados de resistência à compressão com 17 intervalos de classe

6.3 Distribuições de Frequência e Histogramas

- Variante do histograma:

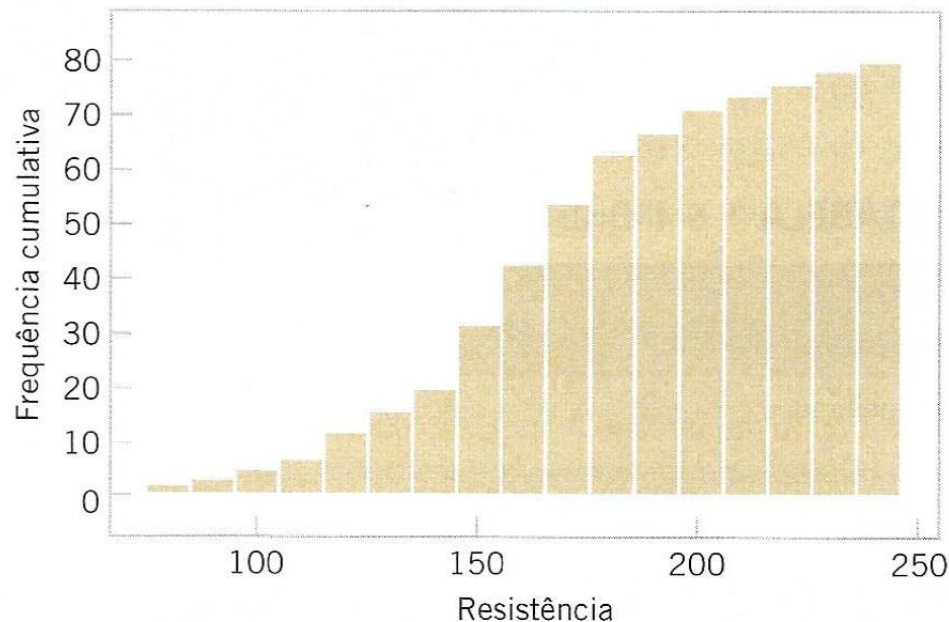


Fig. 6.10 Gráfico de distribuição cumulativa dos dados de resistência à compressão

6.3 Distribuições de Frequência e Histogramas

- O histograma como indicador de forma geral da população:

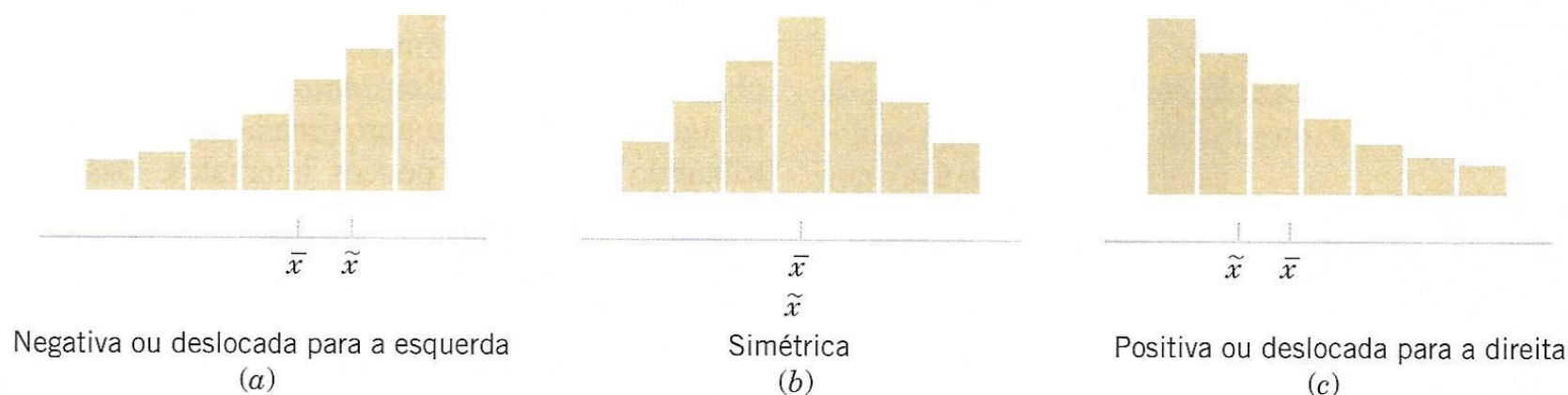


Fig. 6.11 Histograma para distribuições simétricas e deslocadas

6.4 Diagrama de Caixa (*Box plot*)

- O **diagrama de caixa** (*box plot*) é uma representação gráfica que descreve simultaneamente várias características importantes de um conjunto de dados, tais como
 - centro,
 - dispersão,
 - desvio da simetria e
 - identificação das observações que estão surpreendentemente longe da parte principal dos dados:
 - *whisker* (“bigode”);
 - *outlier*;
 - *outlier* extremo.

6.4 Diagrama de Caixa (*Box plot*)

- Exemplo:

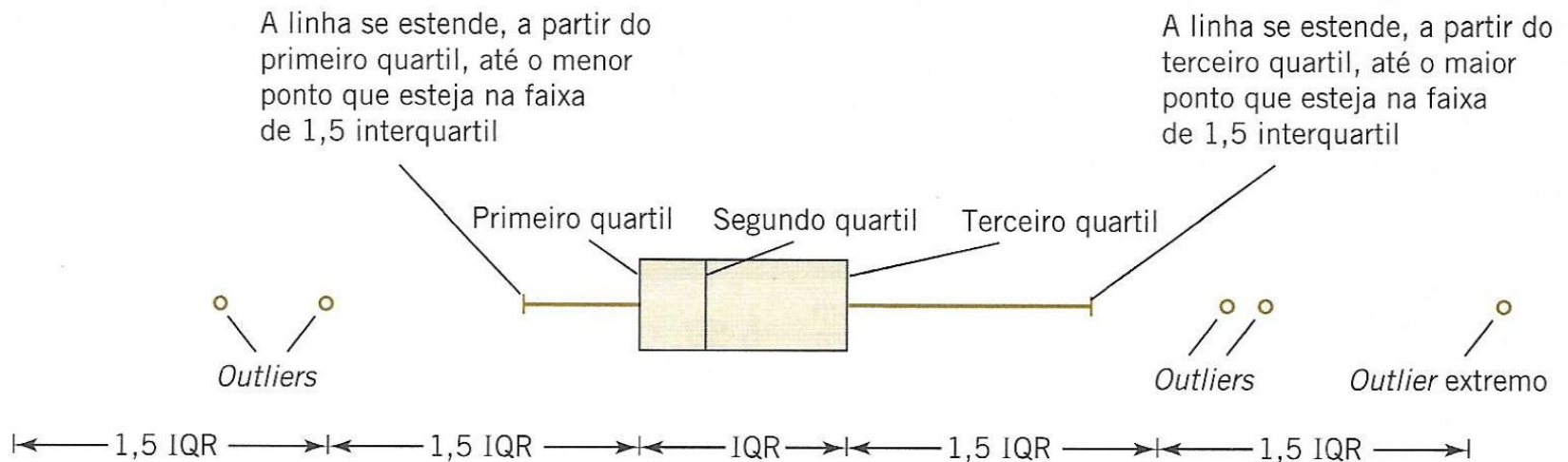


Fig. 6.13 Descrição de um diagrama de caixa

6.4 Diagrama de Caixa (*Box plot*)

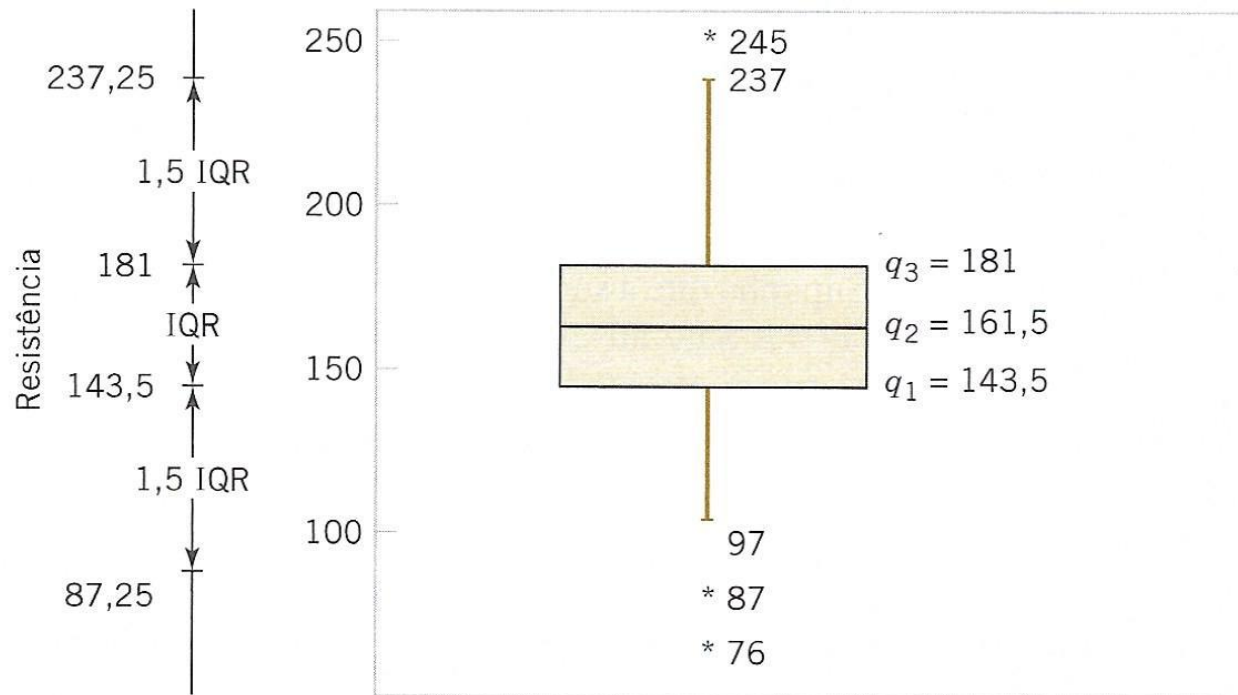
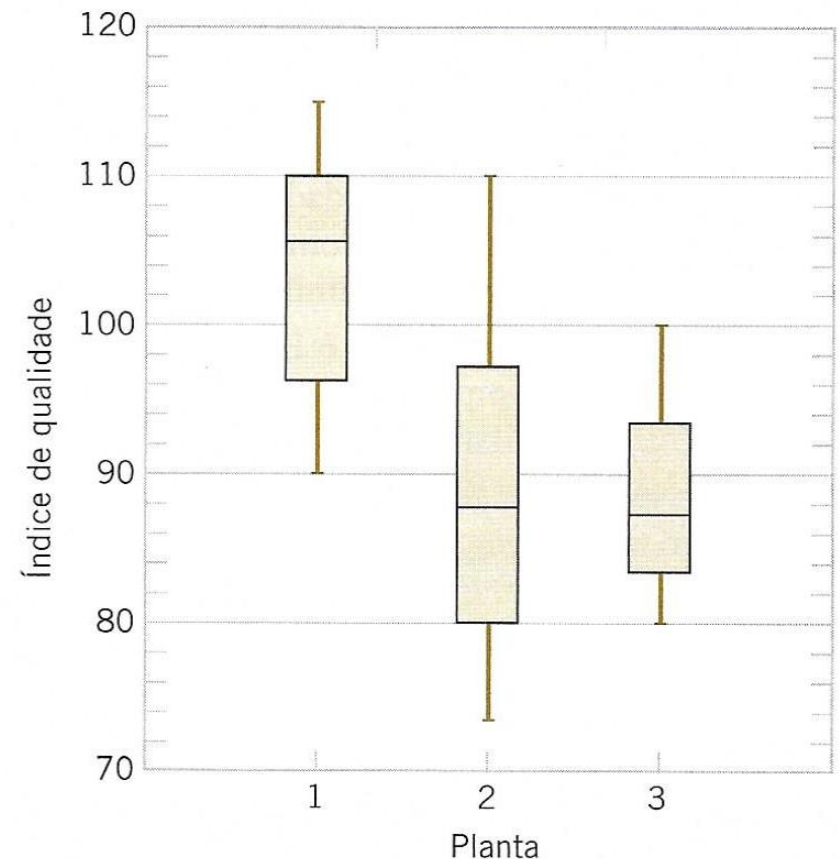


Fig. 6.14 Diagrama de caixa para os dados de resistência à compressão

6.4 Diagrama de Caixa (*Box plot*)

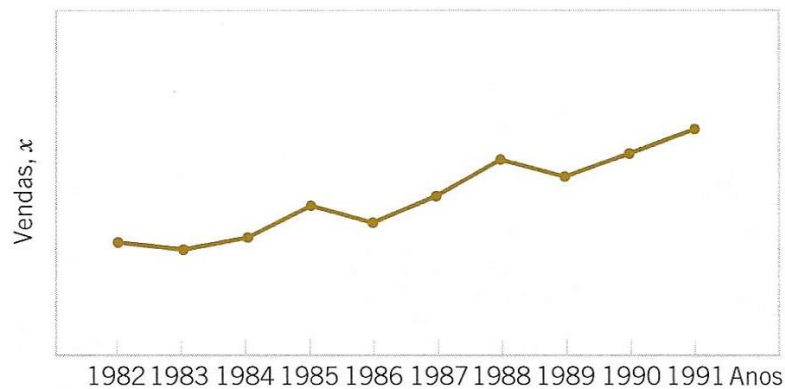
Fig. 6.15 Diagramas de caixa comparativos de um índice de qualidade em três plantas



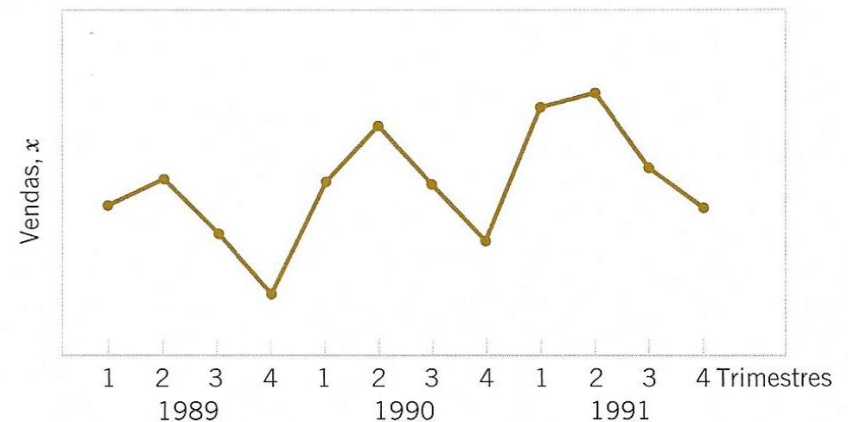
6.5 Gráficos Sequencias Temporais

- Uma **série temporal** ou **sequência temporal** é um conjunto de dados no qual as observações são registradas na ordem em que elas ocorreram;
- Um **gráfico de série temporal** é aquele em que o eixo vertical denota o valor observado da variável (por exemplo, x) e o eixo horizontal denota o tempo (que poderia ser minutos, dias, ano etc.);
- Quando as medidas são grafadas como uma série temporal, frequentemente vemos:
 - **tendências**,
 - **ciclos**, ou
 - **outras características gerais** dos dados que não poderiam ser vistos de outra forma.

6.5 Gráficos Sequencias de Tempo



(a)



(b)

Fig. 6.16 Vendas da companhia por (a) ano e (b) por trimestre

6.5 Gráficos Sequencias de Tempo

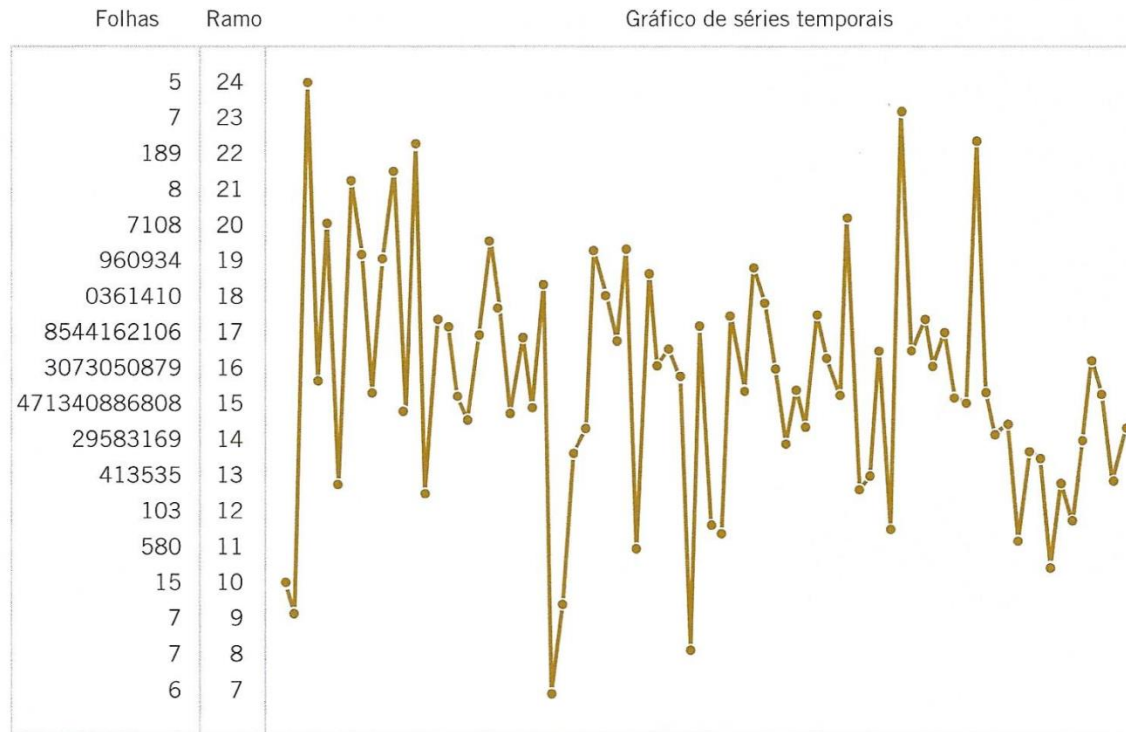


Fig. 6.17 Gráfico digiponto dos dados de resistência à compressão da Tab. 6.2

6.5 Gráficos Sequencias de Tempo

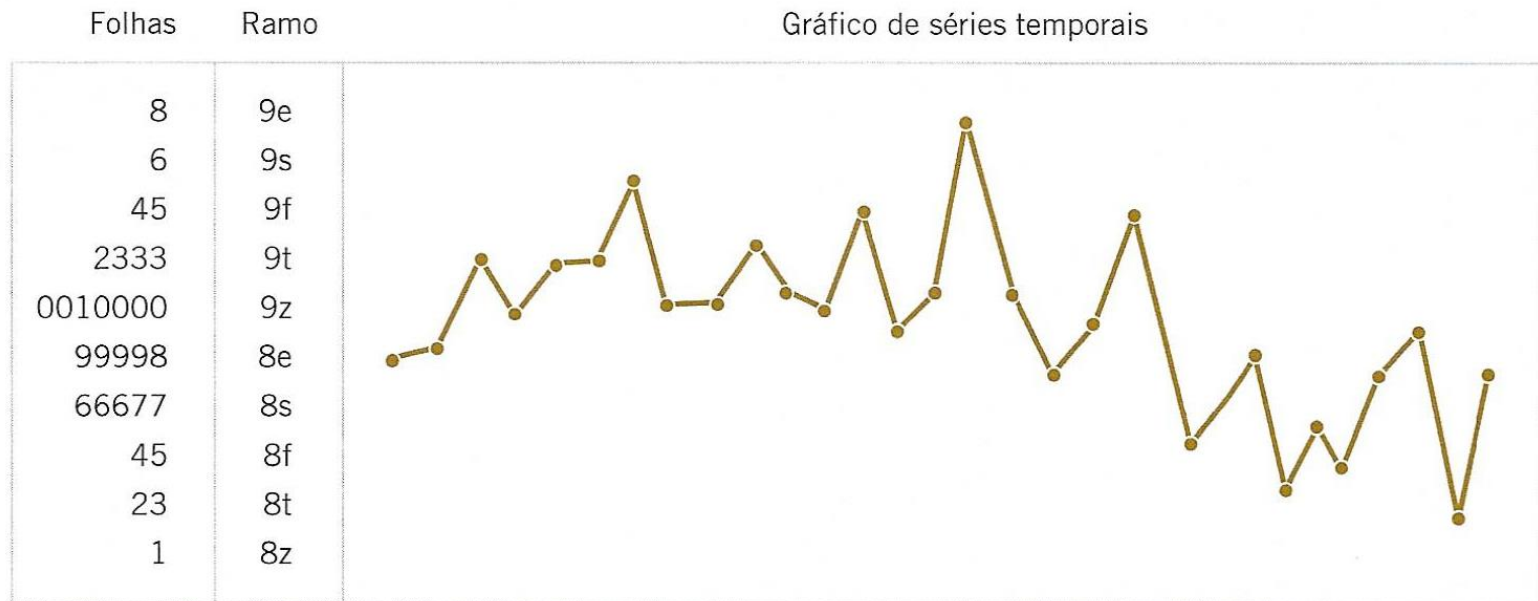


Fig. 6.18 Gráfico digiponto das leituras de concentração de um processo químico, observadas de hora em hora

6.6 Diagramas de Dispersão

- Em muitos problemas, os dados são **multivariados**;
- Diagramas de dispersão são uma ferramenta exploratória **apropriadas** para tais dados;
- Dados apresentados na Tab. 6.5 contém um **exemplo** de dados multivariados;

6.6 Diagramas de Dispersão

Tab. 6.5 Dados sobre a qualidade de vinhos tintos novos

Qualidade	pH	SO ₂ Total	Intensidade de Cor	Cor
19,2	3,85	66	9,35	5,65
18,3	3,75	79	11,15	6,95
17,1	3,88	73	9,40	5,75
15,2	3,66	86	6,40	4,00
14,0	3,47	178	3,60	2,25
13,8	3,75	108	5,80	3,20
12,8	3,92	96	5,00	2,70
17,3	3,97	59	10,25	6,10
16,3	3,76	22	8,20	5,00
16,0	3,98	58	10,15	6,00
15,7	3,75	120	8,80	5,50
15,3	3,77	144	5,60	3,35
14,3	3,76	100	5,55	3,25
14,0	3,76	104	8,70	5,10
13,8	3,90	67	7,41	4,40
12,5	3,80	89	5,35	3,15
11,5	3,65	192	6,35	3,90
14,2	3,60	301	4,25	2,40
17,3	3,86	99	12,85	7,70
15,8	3,93	66	4,90	2,75

6.6 Diagramas de Dispersão

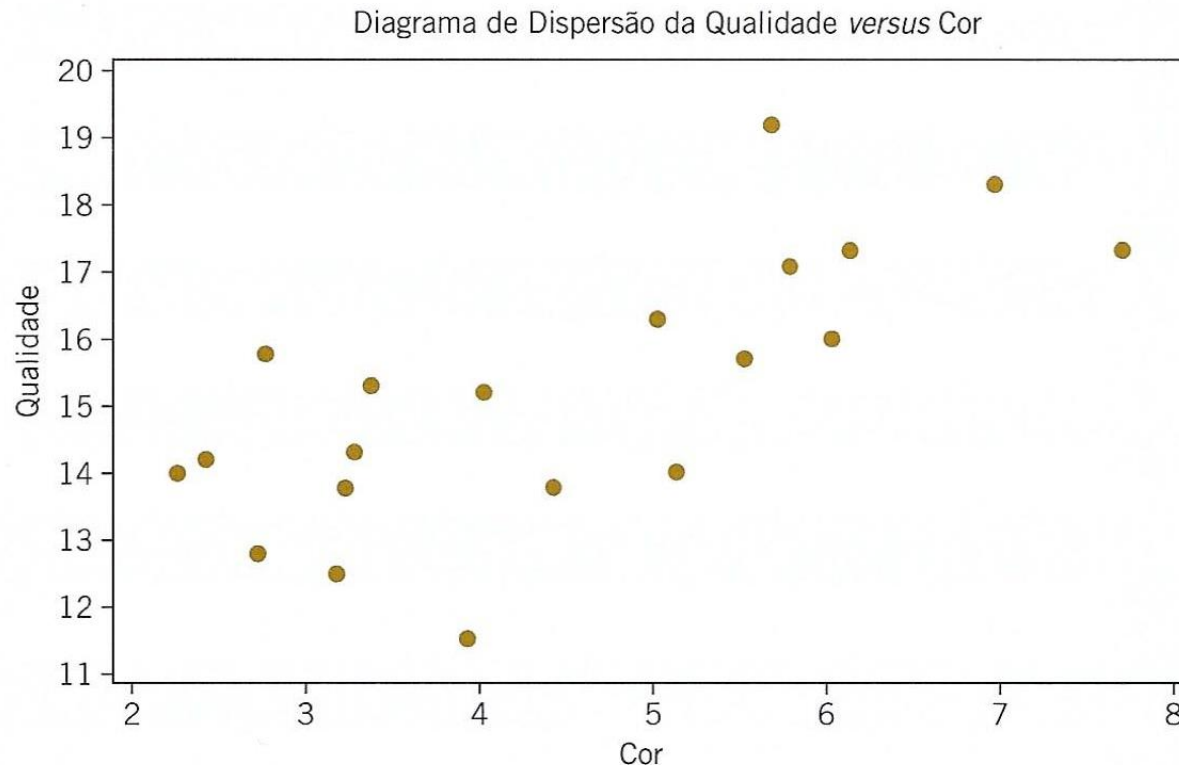


Fig. 6.19 Diagrama de dispersão da qualidade de vinho e da cor a partir da Tab. 6.5

6.6 Diagramas de Dispersão

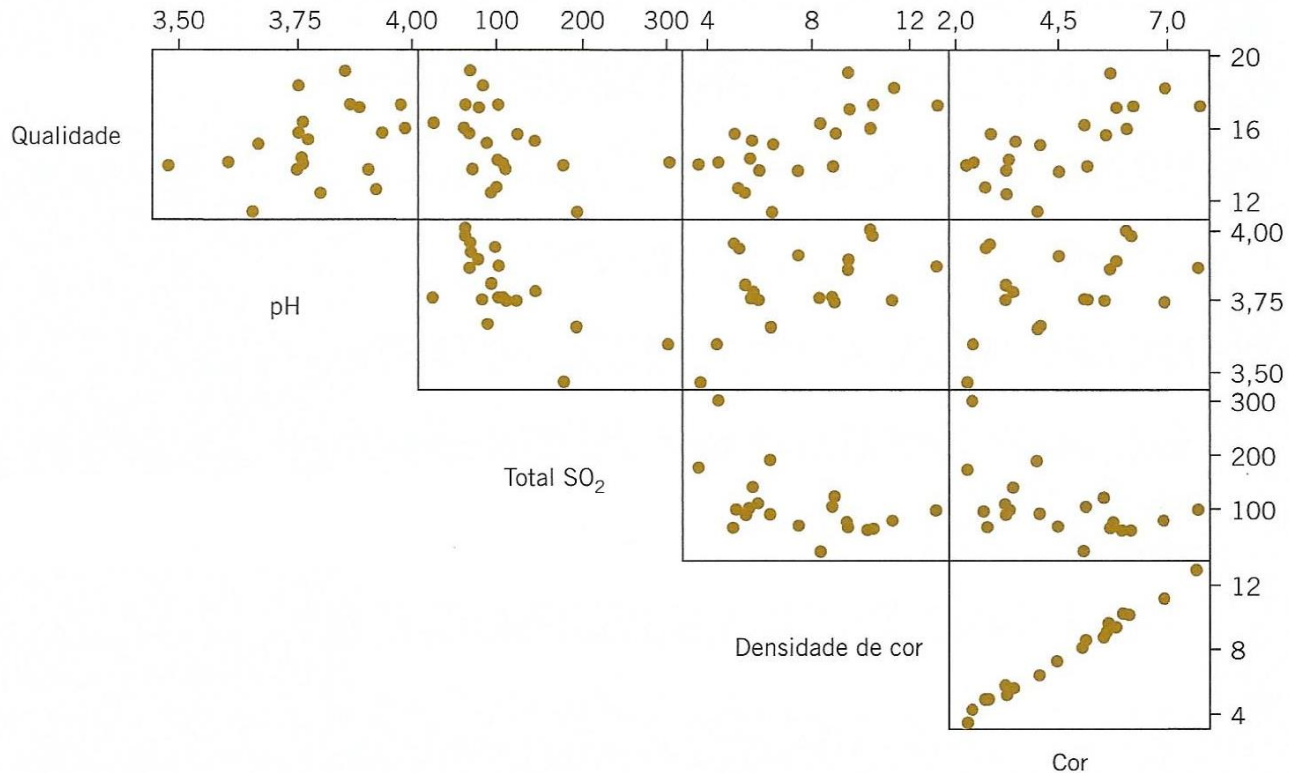
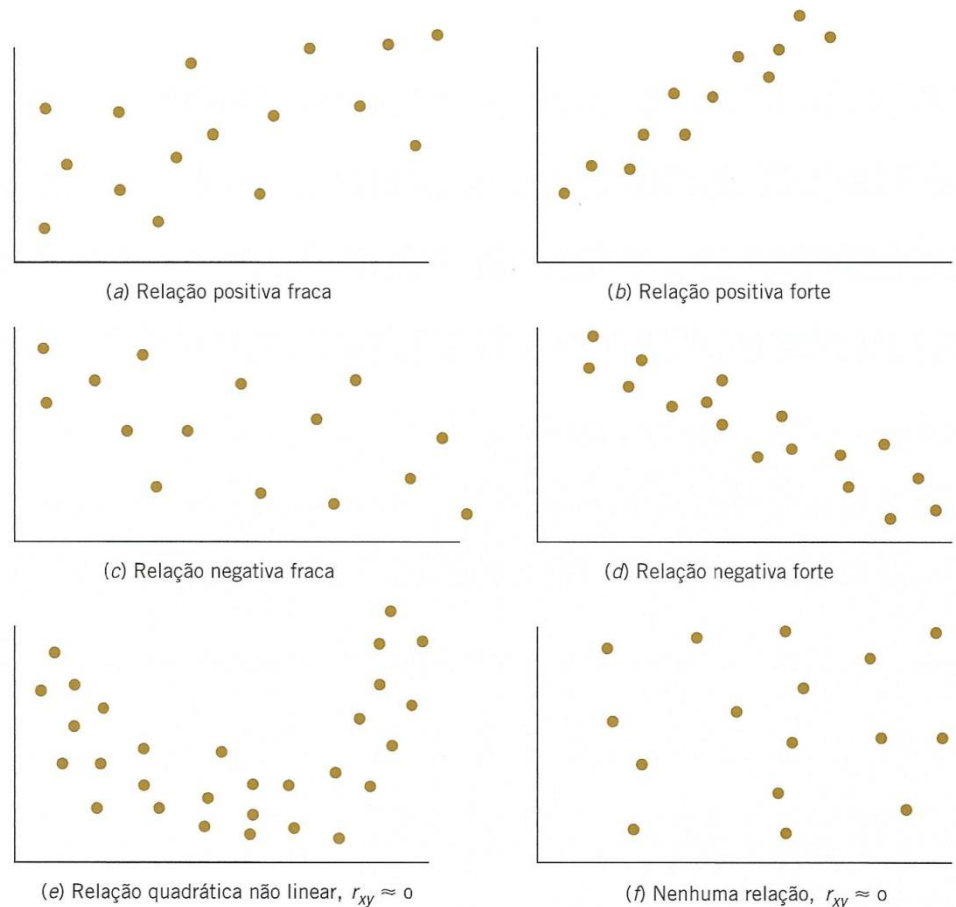


Fig. 6.20 Matriz dos diagramas de dispersão para os dados da qualidade de vinhos da Tab. 6.5

6.6 Diagramas de Dispersão

Fig. 6.21 Relação potencial entre as variáveis



6.7 Gráficos de Probabilidade

- Método para determinar se determinados dados obedecem a uma **distribuição** hipotética (modelo probabilístico hipotético);
- Baseado em análise **visual**;
- Uma distribuição bastante frequente em engenharia é a **Normal**;

6.7 Gráficos de Probabilidade

- Exemplo:**

Vida de Bateria Dez observações sobre o tempo (em minutos) efetivo de vida de serviço de baterias usadas em um computador pessoal são: 176, 191, 214, 220, 205, 192, 201, 190, 183, 185. Imaginemos que a vida da bateria seja modelada adequadamente por uma distribuição normal. Para usar o gráfico de probabilidade de modo a investigar essa hipótese, arranje primeiro as observações em ordem crescente e calcule suas frequências cumulativas $(j - 0,5)/10$ conforme mostrado na Tabela 6-6.

Tab. 6.6 Cálculo para construção de um gráfico de probabilidade normal

j	$x_{(j)}$	$(j - 0,5)/10$	z_j
1	176	0,05	-1,64
2	183	0,15	-1,04
3	185	0,25	-0,67
4	190	0,35	-0,39
5	191	0,45	-0,13
6	192	0,55	0,13
7	201	0,65	0,39
8	205	0,75	0,67
9	214	0,85	1,04
10	220	0,95	1,64

6.7 Gráficos de Probabilidade

- **Exemplo (cont.):**
 - Uma vez que os pontos na Fig. 6.22 caem aproximadamente sobre uma reta, concluímos que a distribuição normal é um modelo apropriado para tais dados;

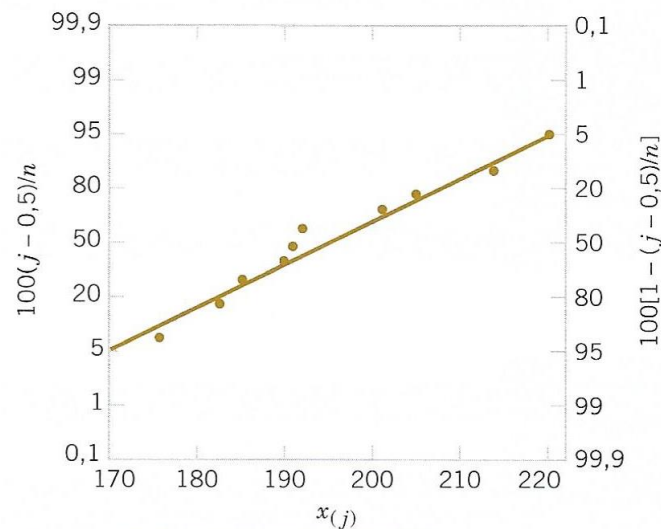
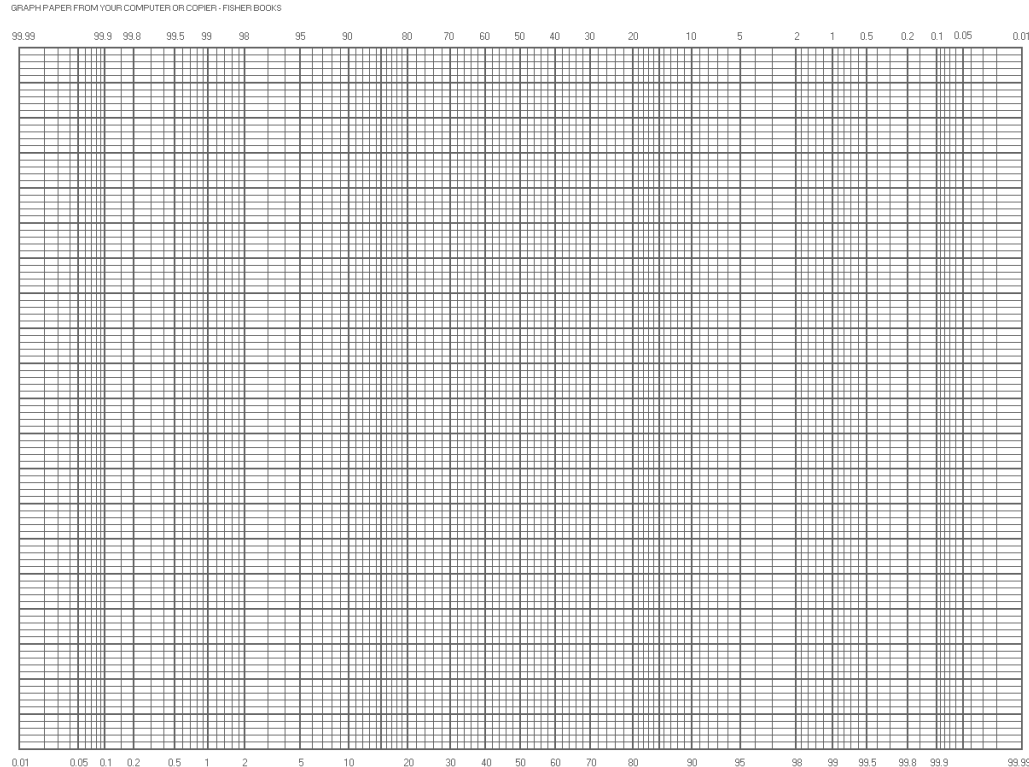


Fig. 6.22 Gráfico de probabilidade normal para a vida da bateria

6.7 Gráficos de Probabilidade

- O papel de probabilidade normal:



TERMOS E CONCEITOS IMPORTANTES

Diagrama de caixa

**Distribuição de
frequência e
histograma**

**Mediana, quartis e
percentis**

Média populacional

**Desvio-padrão
populacional**

**Variância
populacional**

**Distribuição de
frequência relativa**

Média da amostra

**Desvio-padrão da
amostra**

Variância da amostra

**Diagrama de ramo e
folhas**

**Gráficos sequenciais de
tempo**