



# PRO 2271 ESTATÍSTICA I

Prof.Dr. Marco Aurélio de Mesquita  
[marco.mesquita@poli.usp.br](mailto:marco.mesquita@poli.usp.br)

São Paulo, 2008

Escola Politécnica da Universidade de São Paulo | Departamento de Engenharia de Produção





# PRO 2271 ESTATÍSTICA I

**Sumário:**

1. Estatística Descritiva
2. Cálculo de Probabilidades
3. Distribuições de Probabilidades
4. Distribuições Contínuas
5. Intervalos de Confiança
6. Testes de Hipóteses
7. Análise de Variância
8. Correlação e Regressão Linear

Escola Politécnica da Universidade de São Paulo | Departamento de Engenharia de Produção



2



# PRO 2271 ESTATÍSTICA I

## 1. Estatística Descritiva

Escola Politécnica da Universidade de São Paulo | Departamento de Engenharia de Produção



## Introdução

**Estatística:**

- ciência dos dados, proporciona métodos para coleta, representação e análise de dados

**Aplicações:**

- engenharia, administração, economia, ciências sociais, medicina e saúde, meio ambiente etc



**Estatística Descritiva:**

- síntese e representação dos dados

**Inferência Estatística:**

- afirmação sobre a população ou processo em estudo, com base em dados amostrais e sujeitas ao erro estatístico

Escola Politécnica da Universidade de São Paulo | Departamento de Engenharia de Produção 4



## População e Amostra

**População:**

- conjunto de elementos com característica comuns (exemplo, alunos da POLI)

**Amostra:**

- qualquer subconjunto extraído da população (exemplo, uma turma de PRO2271)



**População ou Processo?**

**Variáveis**

- Qualitativas: atributos
- Quantitativas: inteiras (contagem) ou contínuas (medição)

Escola Politécnica da Universidade de São Paulo | Departamento de Engenharia de Produção

5





## Exemplo - Variáveis

Pessoas	Carros
– Nome	– Marca / Modelo
– Idade	– Cor
– Gênero	– Ano de Fabricação
– Formação	– Motor
– Estado Civil	– Potência
– Cargo	– Torque
– Data de Contratação	– Aceleração
– Salário	– Consumo
– Peso	– Autonomia
– Altura	– Comprimento
	– Largura

Escola Politécnica da Universidade de São Paulo | Departamento de Engenharia de Produção

6

## Parâmetros e Estatísticas

**Parâmetro:**

- valores que determinam a distribuição de uma variável aleatória associada a uma população ou processo



**Estatística:**

- valores calculados a partir de uma ou mais amostras, que fornecem estimativa de um parâmetro

Parâmetro	Estatística
$\mu$	$\bar{x}$
$\sigma$	$s$
$p$	$\hat{p}$

Escola Politécnica da Universidade de São Paulo
Departamento de Engenharia de Produção

7

## Amostragem

**Exemplo: Levantamento de Opinião Pública**

- 1. Amostragem Casual**
  - » centenas de milhares de eleitores na Av. Paulista, durante alguns meses de pesquisa, abordados aleatoriamente
- 2. Amostragem Aleatória**
  - » sorteio de dois mil eleitores, a partir do cadastro do TRE
- 3. Amostragem Estratificada**
  - » seleção dos domicílios eleitorais e, em seguida dos eleitores, proporcionalmente à população da região, estado, município, bairro e logradouro

Há diferentes Métodos de Amostragem para cada aplicação específica da Estatística

Escola Politécnica da Universidade de São Paulo
Departamento de Engenharia de Produção

8



## Estatística Descritiva

Apresentar dados amostrais ou populacionais (censo)

Recursos: Tabelas e Gráficos



Relatórios Técnicos, Anuários Estatísticos etc

Exemplos:

- [www.ibge.gov.br](http://www.ibge.gov.br)
- [www.sistemas.usp.br/anuario](http://www.sistemas.usp.br/anuario)
- [www.anfavea.com.br/anuario.htm](http://www.anfavea.com.br/anuario.htm) etc

Escola Politécnica da Universidade de São Paulo | Departamento de Engenharia de Produção

9



## Representação Gráfica

Diagrama de Barras e Circular

Histograma e *Box-Plot*

Diagrama de dispersão

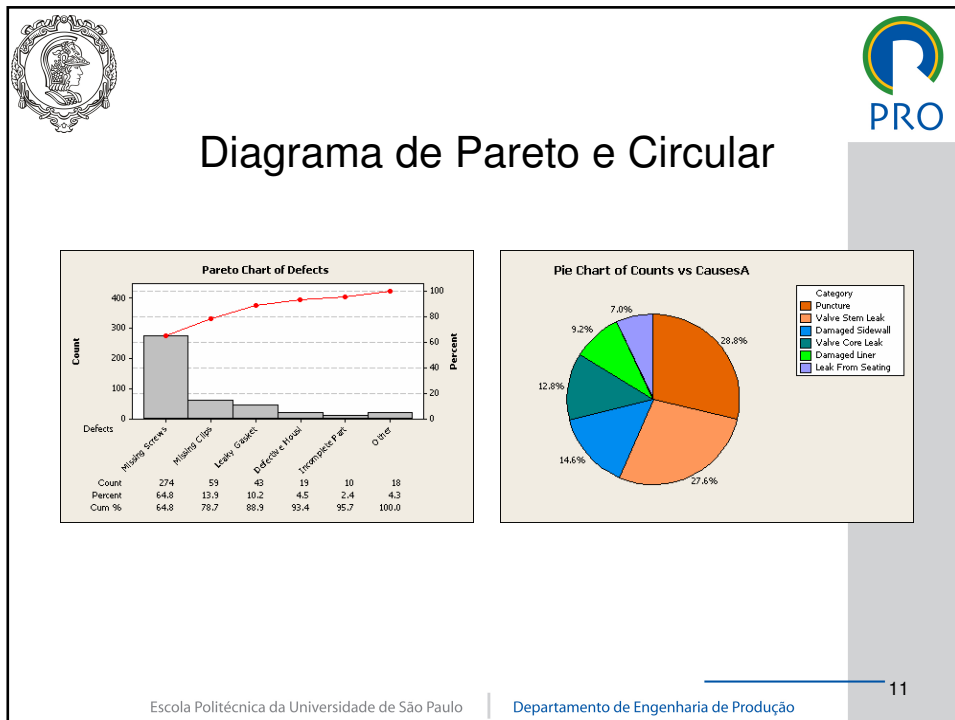
Gráficos de Controle

Séries Temporais

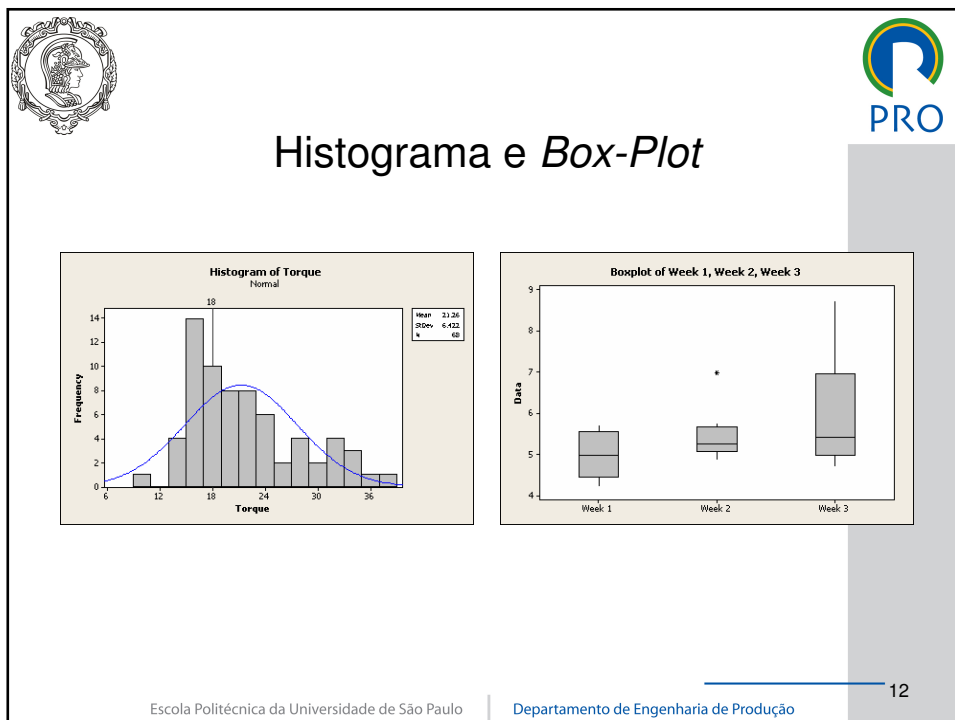
- etc, etc...

Escola Politécnica da Universidade de São Paulo | Departamento de Engenharia de Produção



10



11

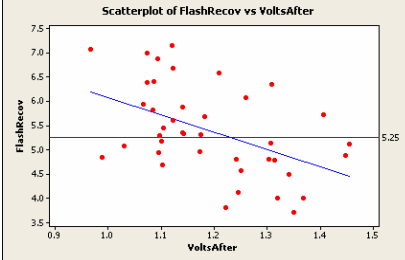


12

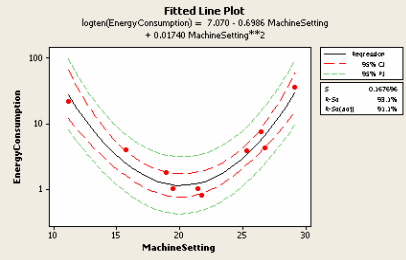
## Diagrama de Dispersão

**Scatterplot of FlashRecov vs VoltsAfter**



**Fitted Line Plot**

logten(EnergyConsumption) = 7.070 - 0.6988 MachineSetting  
+ 0.01740 MachineSetting<sup>2</sup>





S	0.167096
R-Sq	99.1%
R-Sq(Adj)	99.1%

Escola Politécnica da Universidade de São Paulo

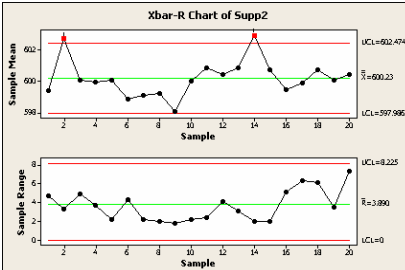
Departamento de Engenharia de Produção

13

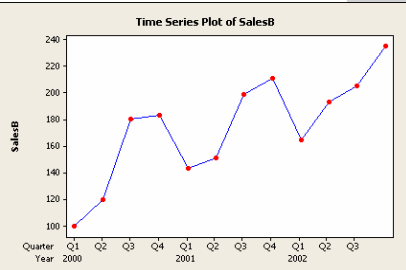



## Gráficos de Controle e Série Temporal

**Xbar-R Chart of SuppZ**





**Time Series Plot of SalesB**



Escola Politécnica da Universidade de São Paulo

Departamento de Engenharia de Produção

14

## Estatísticas

**Variáveis qualitativas**



- Frequência Relativa (%)

**Variáveis quantitativas**

- Medidas de Tendência
  - » média, mediana, quartil etc
- Medidas de Dispersão
  - » amplitude, variância, desvio-padrão etc
- Outros Indicadores
  - » assimetria, curtose etc

15

Escola Politécnica da Universidade de São Paulo | Departamento de Engenharia de Produção

## Exemplo 1 – Variável Qualitativa

**Levantamento de Opinião Pública**


- Amostra: 400 eleitores
- Variável: nível de aprovação
- Contagem: 96 “ótimo / bom”
- Estatística: 24% “ótimo / bom”
- Erro Padrão:  $\pm 2,1\%$

$$n = 400 \quad \hat{p} = 24\% \quad s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 2,1\%$$


16

Escola Politécnica da Universidade de São Paulo | Departamento de Engenharia de Produção





## Exemplo 2 – Variável Quantitativa




77,6	81,2	76,3	79,5	81,9	80,2	84,6	78,5	83,3	83,6
83,2	81,2	79,6	81,9	82,0	82,5	78,3	82,1	81,5	80,0
84,5	83,2	78,4	84,6	80,6	81,3	77,2	83,1	79,0	80,4
81,1	84,7	81,8	80,7	82,1	81,6	80,4	82,4	81,9	78,0
88,1	81,1	<b>75,5</b>	76,3	80,1	85,7	83,0	80,5	79,5	85,1
83,0	78,5	80,4	81,5	79,0	78,7	83,9	84,4	86,9	82,1
81,6	83,7	79,7	87,8	84,6	83,8	79,6	81,6	79,4	82,5
85,4	83,6	82,1	85,0	88,9	81,3	85,9	81,7	78,2	82,7
86,1	80,9	80,8	80,4	81,0	83,9	<b>90,5</b>	85,7	77,5	84,1
79,5	85,9	81,5	85,0	81,5	84,2	86,2	81,4	83,8	77,1


Escola Politécnica da Universidade de São Paulo

Departamento de Engenharia de Produção

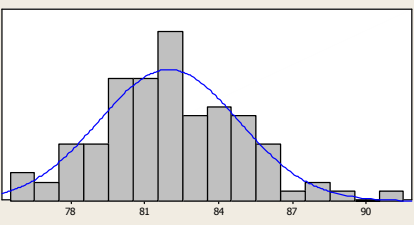
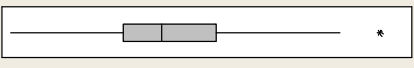
17



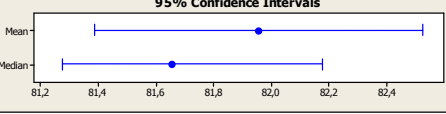
## Exemplo 2 (cont.) – Saída Minitab®



Summary for X

95% Confidence Intervals



Anderson-Darling Normality Test

A-Squared 0,24  
P-Value 0,763

Mean 81,954  
StDev 2,858  
Variance 8,168  
Skewness 0,287284  
Kurtosis 0,192293  
N 100

Minimum 75,500  
1st Quartile 80,125  
Median 81,650  
3rd Quartile 83,875  
Maximum 90,500

95% Confidence Interval for Mean  
81,387 82,521



95% Confidence Interval for Median  
81,274 82,177

95% Confidence Interval for StDev  
2,509 3,320

Escola Politécnica da Universidade de São Paulo

Departamento de Engenharia de Produção

18

## Média amostral

Média aritmética dos valores da amostra



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

Exemplo: 40, 50, 60, 80 e 120

$$\bar{x} = \frac{40 + 50 + 60 + 80 + 120}{5} = \frac{350}{5} = 70$$

19

Escola Politécnica da Universidade de São Paulo | Departamento de Engenharia de Produção

## Mediana

Mediana: valor intermediário da amostra ordenada

- se n ímpar, a mediana será o valor intermediário;
- senão, tomar a média dos dois valores intermediários



Exemplo:

- 40, 50, 60, 80 e 120
- mediana:  $\tilde{x} = 60$

Média ou Mediana, qual o melhor indicador?

20

Escola Politécnica da Universidade de São Paulo | Departamento de Engenharia de Produção

## Quartil

Quartis: dividem a amostra ordenada em 4 grupos

- 1o Quartil: 25% menores
- 2o Quartil = Mediana
- 3o Quartil: 25% maiores



Exemplo: notas da primeira prova de uma turma de 40 alunos

- n=40:  $x_{\min}=0$   $Q_1=4,5$   $Md=5,5$   $Q_3=7,0$   $x_{\max}=9,5$

Outros indicadores: Decil (10%) e Percentil (1%)

21

Escola Politécnica da Universidade de São Paulo | Departamento de Engenharia de Produção

## Medidas de Dispersão


Amplitude:  $R = x_{\max} - x_{\min}$  (*Range*)

Variância: 
$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$


Desvio Padrão:  $s = \sqrt{s^2}$  (*standard deviation*)

22

Escola Politécnica da Universidade de São Paulo | Departamento de Engenharia de Produção



## Cálculo da Variância



Exemplo 1:    40    45    50    55    60

$$s^2 = \frac{(40-50)^2 + (45-50)^2 + (50-50)^2 + (55-50)^2 + (60-50)^2}{5-1}$$

$$s^2 = 62,5 \Rightarrow s = \sqrt{s^2} = 7,9$$


Exemplo 2:    10    30    50    70    90

$\bar{x} = 50 \quad s = 31,6$


Escola Politécnica da Universidade de São Paulo

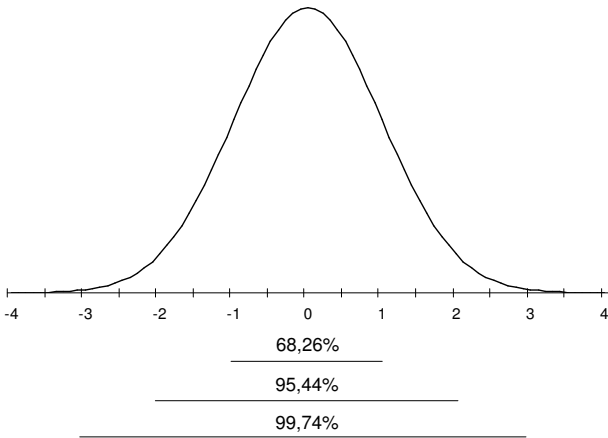
Departamento de Engenharia de Produção

23



## Desvio Padrão - Curva Normal







Escola Politécnica da Universidade de São Paulo

Departamento de Engenharia de Produção

24

## Cálculo da Variância (cont.)

Fórmula Alternativa de Cálculo:



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

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

onde: T = soma simples e  
Q = soma dos quadrados

25

Escola Politécnica da Universidade de São Paulo | Departamento de Engenharia de Produção

## Cálculo da Variância (cont.)

Exemplo: 40, 45, 50, 55 e 60

$$n = 5 \quad T = 250$$



$$Q = 40^2 + 45^2 + 50^2 + 55^2 + 60^2 = 12.750$$

$$\Rightarrow s^2 = \frac{12750 - 250^2/5}{5-1} = 62,5$$

$$s^2 = \frac{Q - \frac{T^2}{n}}{n-1}$$

26

Escola Politécnica da Universidade de São Paulo | Departamento de Engenharia de Produção



## Resumo Estatístico

Amostra: 40 45 50 55 60

Estatística	Valor
Tamanho da Amostra	5
Média	50,0
Mediana	50
Amplitude	20
Variância	62,5
Desvio-padrão	7,9
Coef. de Variação	15,8%

27

Escola Politécnica da Universidade de São Paulo | Departamento de Engenharia de Produção

## Propriedades da Média e Variância

Seja  $x_1, x_2, \dots, x_n$  uma amostra aleatória e  $y=ax+b$ , então:

1.  $\bar{y} = a \cdot \bar{x} + b$
2.  $s_y^2 = a^2 \cdot s_x^2 \quad (s_y = |a| \cdot s_x)$


Exemplo: 40, 45, 50, 55 e 60,  $y=(x-50)/5 \Rightarrow x=5y+50$

$y: -2, -1, 0, +1 \text{ e } +2 \Rightarrow \bar{y} = 0 \quad s_y^2 = 2,5$

$\Rightarrow \bar{x} = 50 \quad s_x^2 = 62,5$


28

Escola Politécnica da Universidade de São Paulo | Departamento de Engenharia de Produção



## Cálculo da Média e Variância

### a partir da Distribuição de Freqüências



Média:  $\bar{x} = \frac{\sum_{i=1}^m x_i \cdot f_i}{n} = \frac{T}{n}$


Variância:

$$s^2 = \frac{\sum_{i=1}^m x_i^2 \cdot f_i - \frac{(\sum_{i=1}^m x_i \cdot f_i)^2}{n}}{n-1} = \frac{Q - \frac{T^2}{n}}{n-1}$$


Escola Politécnica da Universidade de São Paulo

Departamento de Engenharia de Produção

29

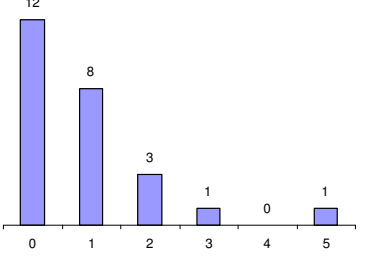


## Exemplo



Determine a média e o desvio padrão da amostra abaixo.

$x_i$	$f_i$	$x_i f_i$	$x_i^2 f_i$
0	12	0	0
1	8	8	8
2	3	6	12
3	1	3	9
4	0	0	0
5	1	5	25
$n = 25$		$T = 22$	$Q = 54$



$\bar{x} = 0,88$        $s = 1,20$

Escola Politécnica da Universidade de São Paulo

Departamento de Engenharia de Produção

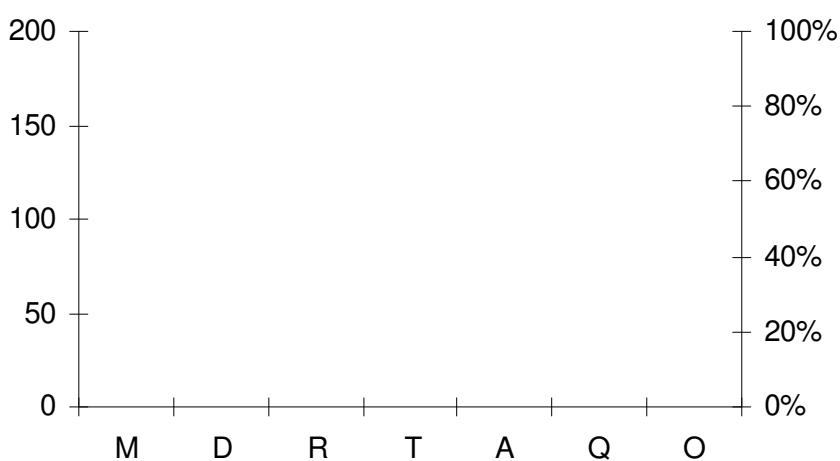
30

EXERCÍCIOS

- 1) Um processo de fabricação é controlado, retirando-se regularmente amostras com 250 itens. Após a inspeção de 20 amostras, obteve-se a seguinte distribuição de defeitos: Mancha, 104; Trinca, 10; Deformação, 42; Quebrado, 4; Risco, 20; Aspereza, 6 e 14 outros defeitos. Construa um **gráfico de Pareto** e analise os resultados. Qual a probabilidade de uma peça manchada?

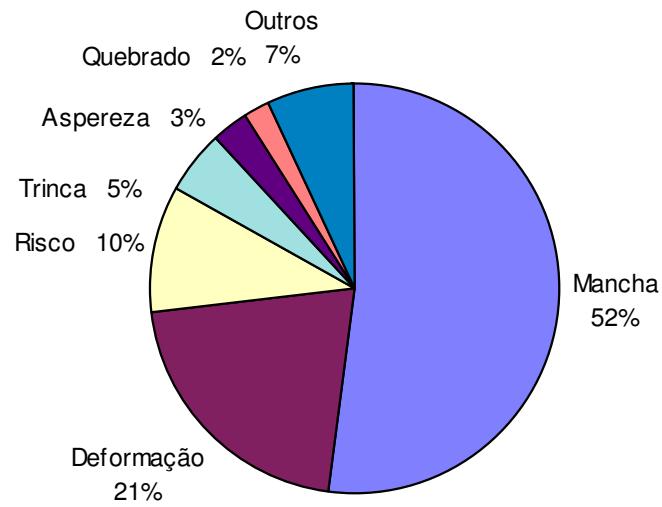
**Tabela 1.1** Distribuição de freqüências dos defeitos no processo de inspeção.

Tipo de Defeito	Freqüência	Freqüência Acumulada	Freqüência Relativa	Freq. Relativa Acumulada
Mancha				
Deformação				
Risco				
Trinca				
Aspereza				
Quebrado				
Outros				
Total				

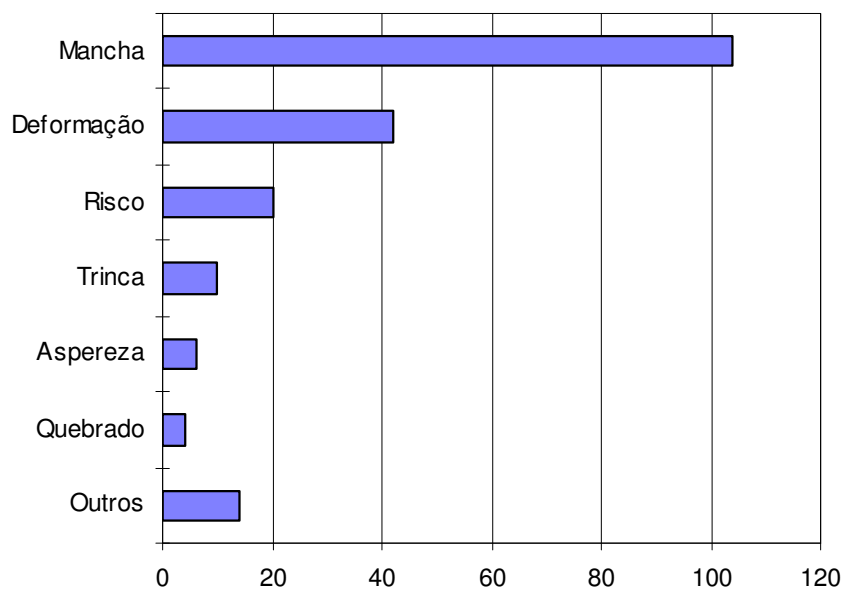


**Figura 1.1** Gráfico de Pareto por tipo de defeito.





**Figura 1.2** Diagrama Circular por tipo de defeito.

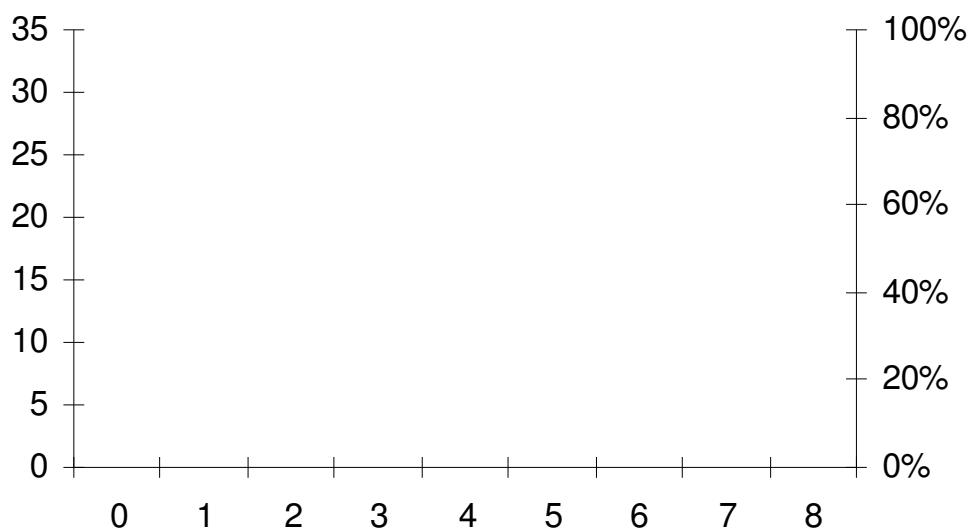


**Figura 1.3** Diagrama de Barras por tipo de defeito.

- 2) Os dados abaixo correspondem à quantidade de defeitos por unidade, em uma amostra com 100 dispositivos. Qual o tipo de variável em questão? Represente graficamente os dados da amostra e determine o número médio de defeitos por unidade.

**Tabela 1.2** Distribuição de defeitos por unidade.

Quantidade de defeitos	Frequência	Freq. Rel. Acumulada	
0	20		
1	30		
2	25		
3	14		
4	7		
5	3		
6	0		
7	1		
8	0		
Total	100		



**Figura 1.4** Distribuição de defeitos por unidade.

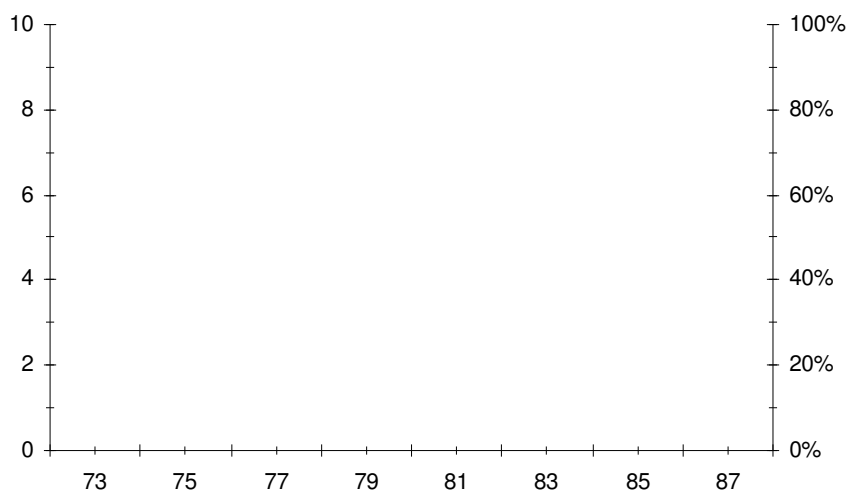
- 3) Os dados abaixo representam a carga estática máxima em um ensaio de resistência. Elabore um diagrama de ramos-e-folhas a partir desta da amostra. A seguir, determine uma distribuição de freqüências e construa o histograma correspondente.

76,3	81,5	78,2	77,9	80,2
82,6	78,0	76,7	75,9	78,5
78,4	79,4	79,0	81,6	74,6
77,6	79,6	81,1	77,3	80,2
78,5	84,1	81,5	82,5	79,1

74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84

**Tabela 1.3** Distribuição de freqüências ...

Classe	Freqüência Absoluta	Freq. Abs. Acumulada	Freqüência Relativa	Freq. Rel. Acumulada
74,0 - 75,9				
76,0 - 77,9				
78,0 - 79,9				
80,0 - 81,9				
82,0 - 83,9				
84,0 - 85,9				
Total				



**Figura 1.5** Distribuição de Freqüências ...

**Descriptive Statistics:**

Variable	N	Mean	Median	TrMean	StDev	SE Mean
X	25	79,212	79,000	79,200	2,303	0,461

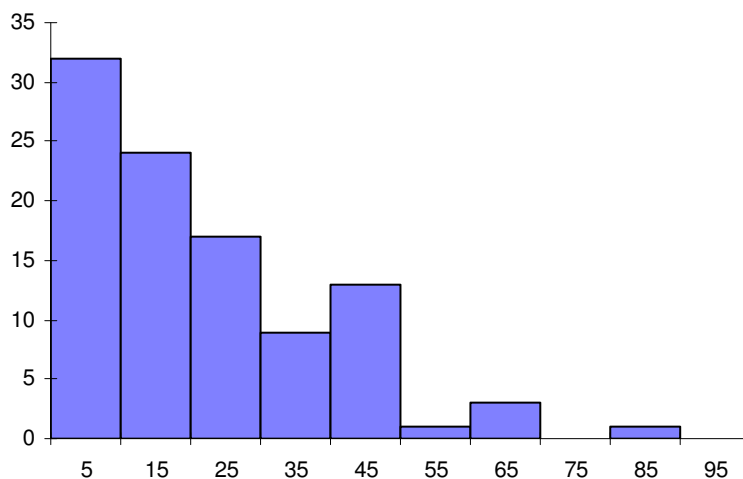
Variable	Minimum	Maximum	Q1	Q3
X	74,600	84,100	77,750	81,300

- 4) Os dados abaixo constituem os tempos de reparo (em min) de um equipamento. Determine a distribuição de freqüências e construa o histograma correspondente. ■

43	32	40	08	03	22	46	01	20	16
19	48	08	06	60	18	49	09	16	10
09	66	05	14	12	26	09	19	04	04
81	45	29	22	36	01	06	30	36	06
28	20	01	18	44	55	66	10	03	13
35	34	01	04	11	30	20	27	14	20
07	10	29	10	40	18	40	01	46	05
15	48	27	08	31	01	17	13	04	21
10	03	34	09	23	47	47	05	25	28
01	14	01	17	12	24	07	02	01	11

**Tabela 1.4** Distribuição dos tempos de reparo.

Classe	Ponto Médio	Freqüência	Freqüência Acumulada
0 - 10	5	32	32
10 - 20	15	24	56
20 - 30	25	17	73
30 - 40	35	9	82
40 - 50	45	13	95
50 - 60	55	1	96
60 - 70	65	3	99
70 - 80	75	0	99
80 - 90	85	1	100
Total		100	



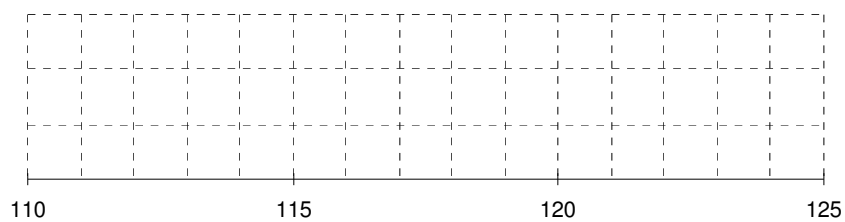
$$\bar{x} = 21,0$$

$$s = 17,4$$

**Figura 1.6** Distribuição dos tempos de reparo.

- 5) Calcule a média, a mediana, os quartis, a amplitude e o desvio padrão da amostra de valores abaixo. A seguir, represente os dados graficamente.

112,7   113,6   114,2   114,4   114,5   115,3   115,4   115,8  
116,2   116,8   116,9   117,0   117,1   117,7   118,2   123,6



**Figura 1.7** Diagrama *Box Plot*.

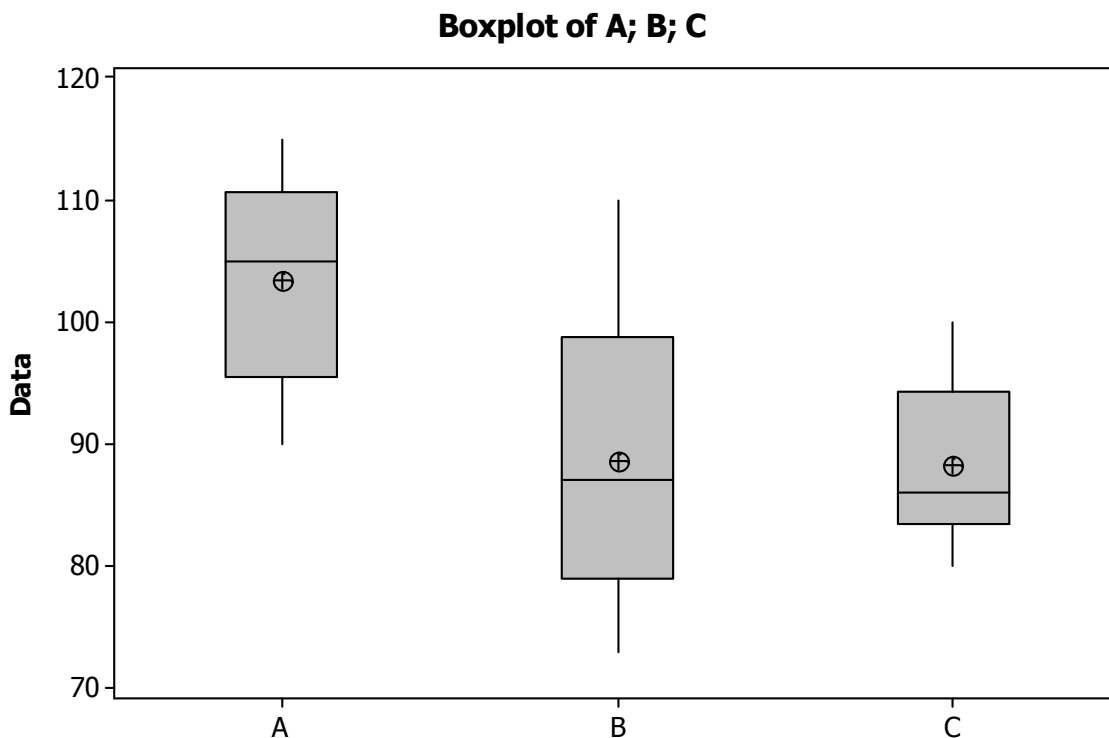
- 6) Os dados abaixo representam o desempenho de três grupos de 10 alunos em um teste. Construa um *box plot* para cada amostra e compare os resultados.

**Tabela 1.5** Resultados dos testes de cada grupo.

Grupo	Desempenhos									
1	90	94	96	100	104	106	106	110	113	115
2	73	76	80	82	86	88	91	97	104	110
3	80	82	84	86	86	86	87	93	98	100

**Descriptive Statistics:**

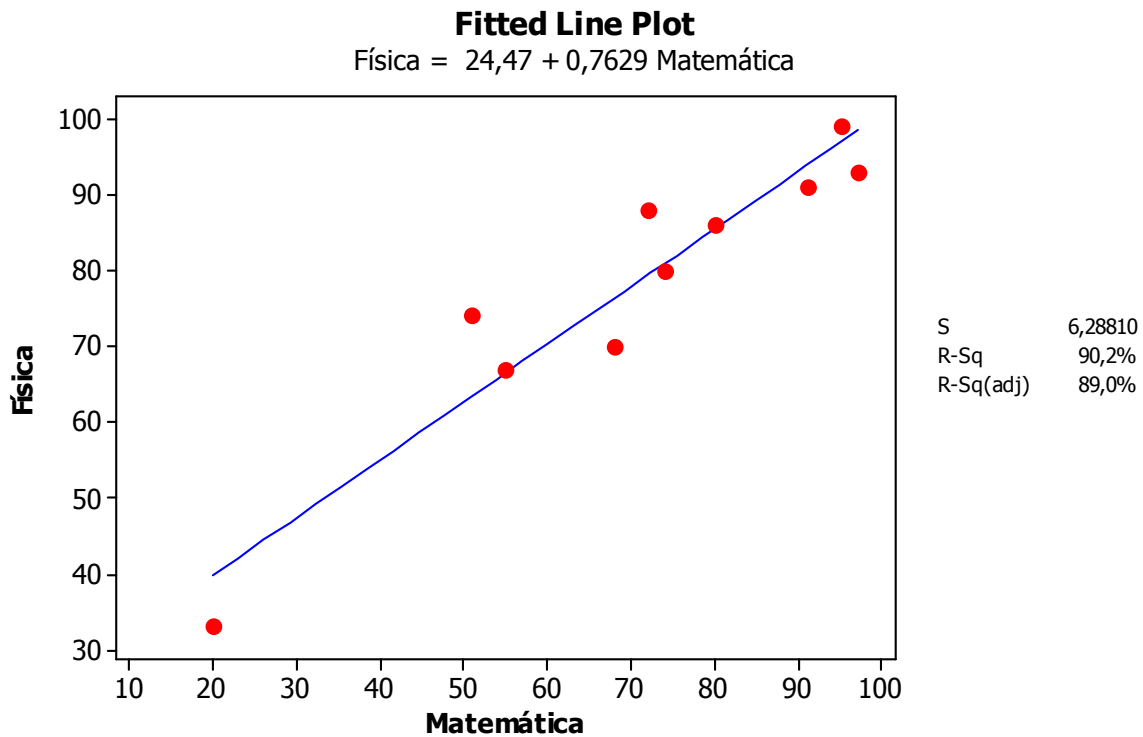
Variable	n	Mean	StDev	Minimum	Maximum	Q1	Median	Q3
A	10	103,40	8,29	90,00	115,00	95,50	105,00	110,75
B	10	88,70	12,01	73,00	110,00	79,00	87,00	98,75
C	10	88,20	6,65	80,00	100,00	83,50	86,00	94,25



**Figura 1.8** Diagrama comparativo dos desempenhos.

7) Os dados abaixo representam as notas finais de doze alunos nas disciplinas de Matemática e Física. Construa um **diagrama de dispersão** e verifique se há correlação linear entre as notas.

Matemática	Física
51	74
68	70
72	88
97	93
55	67
95	99
20	33
91	91
74	80
80	86



**Figura 1.9** Diagrama de Dispersão entre as notas de Matemática e Física.



EXERCÍCIOS PROPOSTOS

- 1) Na inspeção de um lote de placas de circuito impresso, foram detectadas as seguintes desconformidades: componente incorreto, 210; soldagem incorreta, 121; faltando componente, 103; falha de componente, 68; outros, 81. Após a implementação de ações de melhoria, a incidência de defeitos em um lote com o mesmo número de placas resultou em: componente incorreto, 63; soldagem incorreta, 100; faltando componente, 24; falha de componente, 72; outros, 58. Represente graficamente os dados e compare as situações "antes" e "depois".
  
- 2) Determine uma distribuição de freqüências e construa o histograma correspondente para os dados abaixo. Calcule a média, os quartis, o desvio padrão amostrais e a porcentagem de valores fora do intervalo  $82,5 \pm 2,5$ .

75,5	78,4	79,6	80,5	81,3	81,7	82,4	83,6	84,5	85,7
76,3	78,5	79,6	80,6	81,3	81,8	82,5	83,6	84,6	85,9
76,3	78,5	79,7	80,7	81,4	81,9	82,5	83,7	84,6	85,9
77,1	78,7	80,0	80,8	81,5	81,9	82,7	83,8	84,6	86,1
77,2	79,0	80,1	80,9	81,5	81,9	83,0	83,8	84,7	86,2
77,5	79,0	80,2	81,0	81,5	82,0	83,0	83,9	85,0	86,9
77,6	79,4	80,4	81,1	81,5	82,1	83,1	83,9	85,0	87,8
78,0	79,5	80,4	81,1	81,6	82,1	83,2	84,1	85,1	88,1
78,2	79,5	80,4	81,2	81,6	82,1	83,2	84,2	85,4	88,9
78,3	79,5	80,4	81,2	81,6	82,1	83,3	84,4	85,7	90,5

- 3) Considerem-se as amostras I e II de um processo, apresentadas abaixo.  
 Pede-se:
  - a) construa um *box-plot* para cada uma das amostras;
  - b) calcule a média e o desvio padrão de cada amostra;
  - c) compare os resultados obtidos.

I					II				
49,0	49,3	49,5	49,9	50,2	51,7	51,6	52,0	52,3	52,6
49,2	49,3	49,6	50,0	50,2	50,4	51,6	52,1	52,4	52,6
49,2	49,4	49,7	50,2	50,5	51,1	51,7	52,3	52,5	52,8
49,2	49,5	49,7	50,2	50,7	51,3	51,9	52,3	52,5	52,9
					51,5	52,0	52,3	52,5	52,9

- 4) Um experimento para avaliar a durabilidade de certo componente forneceu os seguintes resultados (em horas) : 75, 95, 100+, 81, 93, 66, 47, 100+, 35 e 69. Quais medidas de tendência central poderiam ser utilizadas para a variável em questão ? Como seria possível avaliar a dispersão da variável aleatória em questão. Represente os dados graficamente.

- 5) Calcule a média, a mediana, os quartis e o desvio padrão da seguinte distribuição de freqüências:

$x_i$	$f_i$
75	01
77	06
79	16
81	32
83	22
85	16
87	04
89	02
91	01
	100

- 6) Seja  $x_1, x_2, \dots, x_n$  uma amostra aleatória e  $\bar{x}, s^2, s$ , respectivamente, a média, a variância e o desvio padrão amostrais. Determine a média, variância e desvio padrão de  $z_1, z_2, \dots, z_n$  onde  $z_i = (x_i - \bar{x})/s$ .
- 7) Utilizando a planilha MS Excel, obtenha 25 amostras aleatórias com 10 valores cada a partir de uma distribuição normal com média 50 e desvio padrão 2. Construa os gráficos de controle da média e do desvio padrão.