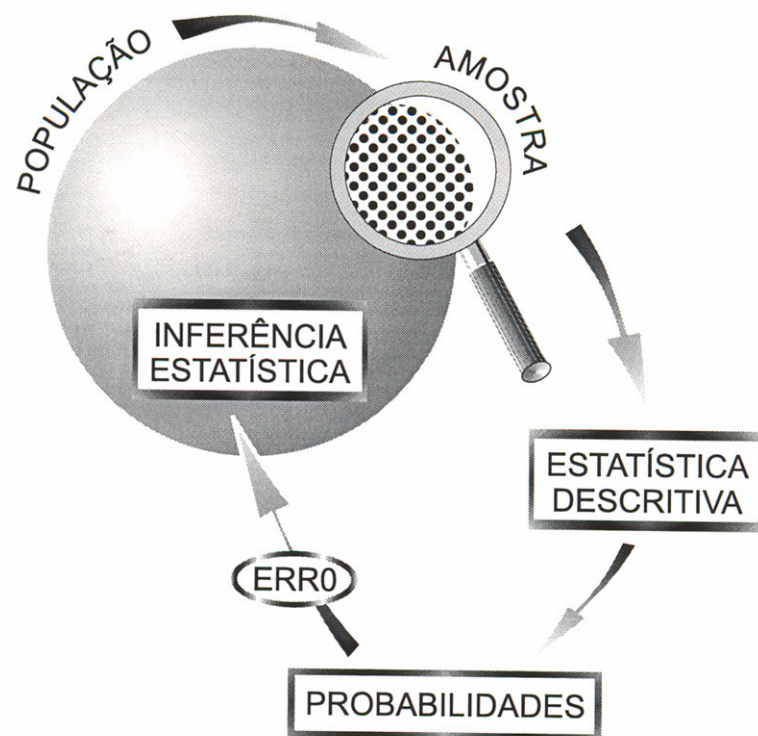


# Estatística Descritiva

Ana Amelia Benedito Silva

*Figura 1.2*  
*Visão sistêmica da*  
*Estatística.*



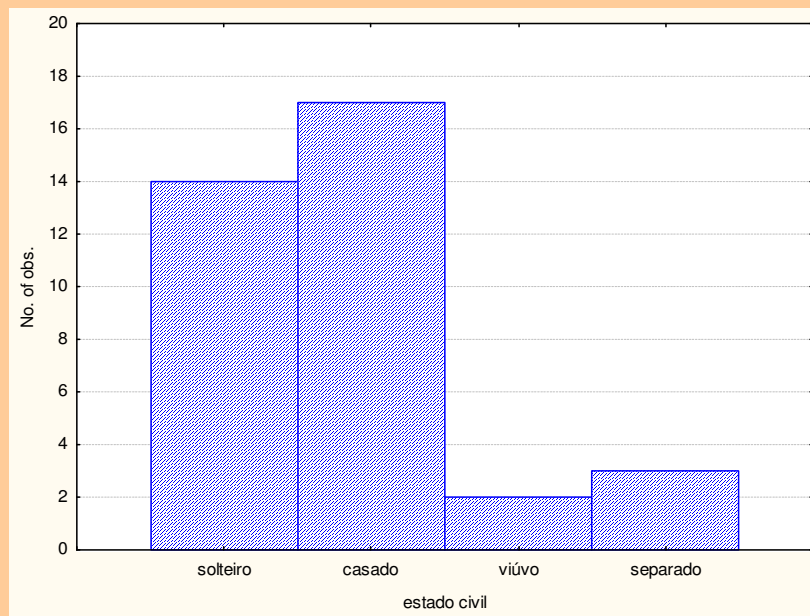
# Representação de dados

- Tabelas
- Gráficos
  - Dados qualitativos
    - Barras (horizontais ou verticais)
    - Setores (pizzas)
  - Dados quantitativos
    - Histograma para dados discretos
    - Histograma para dados contínuos

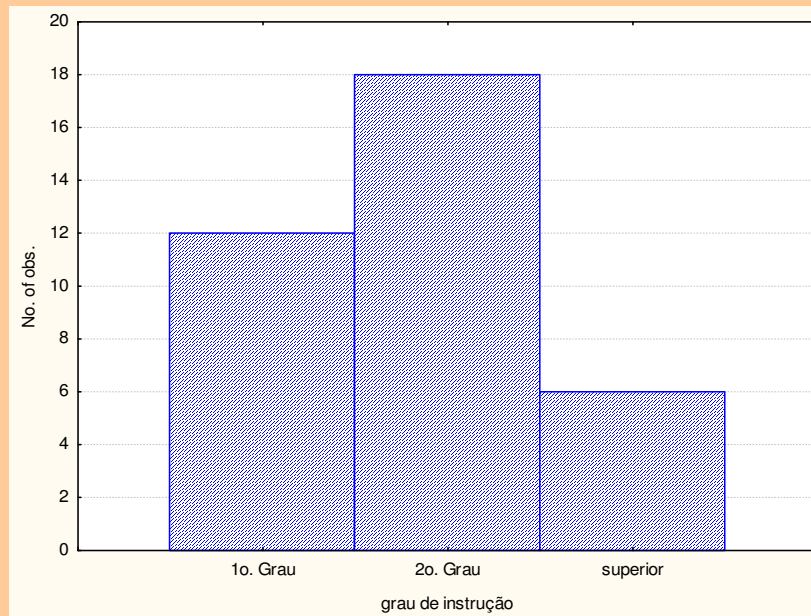
# Dados de 36 trabalhadores de uma empresa

nome	estado civil	escolaridade	Filhos	salário	idade	origem
Natanael	solteiro	1o. Grau	0	4	26	interior
Manoel	casado	1o. Grau	1	4.56	32	capital
.....	.....	.....	.....	.....	...	.....
Walter	casado	2o. Grau	2	15.99	35	capital
Hernando	casado	2o. Grau	2	19.4	48	capital
Gregório	casado	superior	3	23.3	42	interior

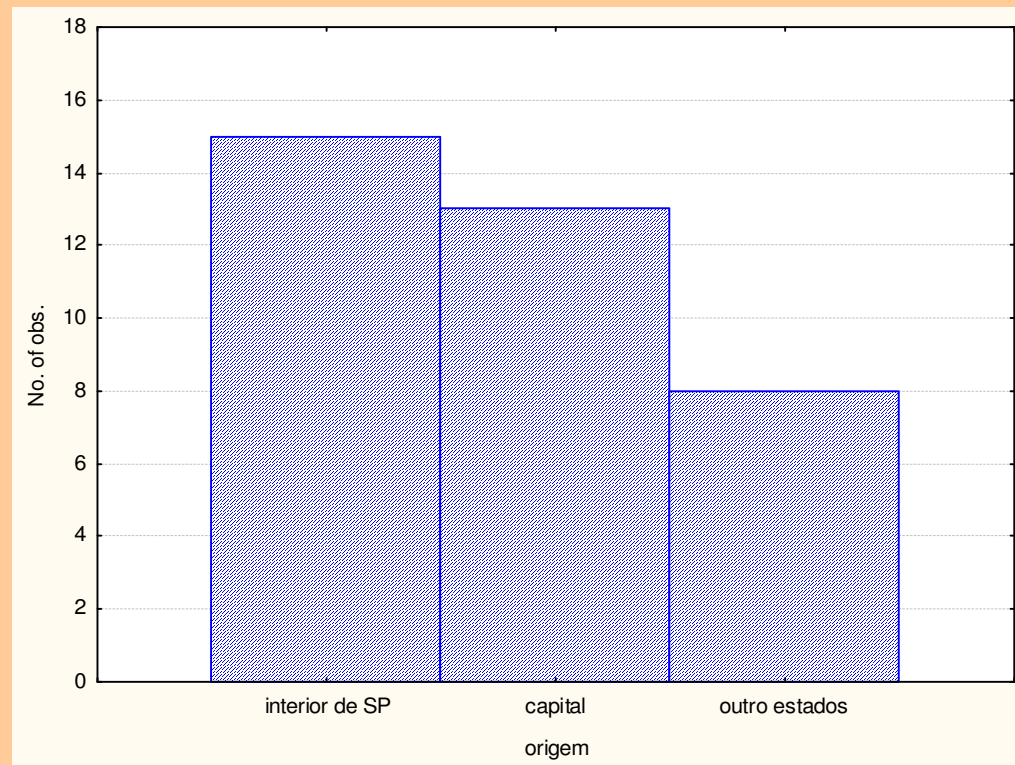
Categorias		
	Count	Percent
solteiro	14	38,9
casado	17	47,2
viúvo	2	5,6
separado	3	8,3



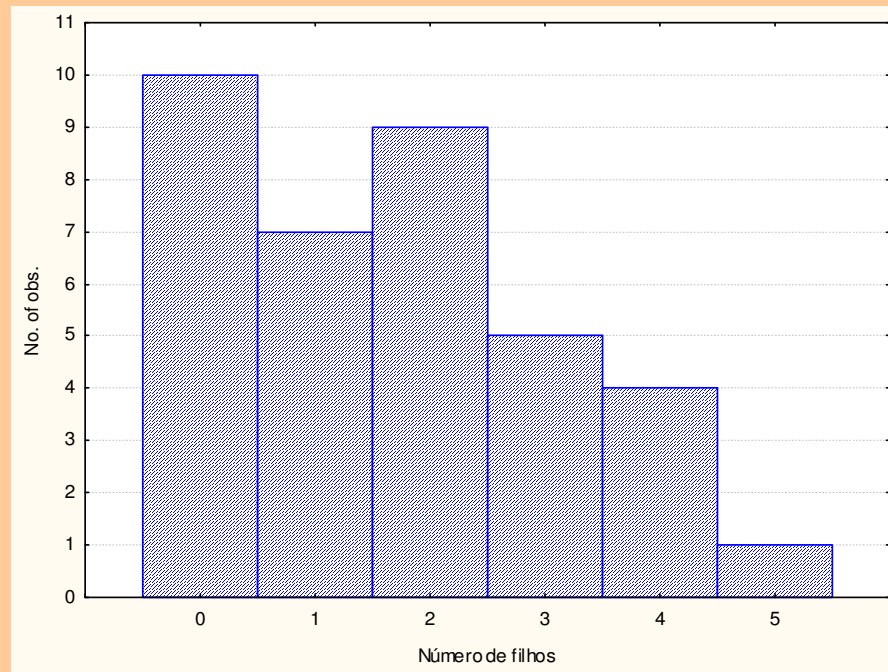
Categorias		
	Count	Percent
1o. Grau	12	33,3
2o. Grau	18	50,0
superior	6	16,7



Categorias		
	Count	Percent
interior de SP	15	41,7
capital	13	36,1
outro estados	8	22,2

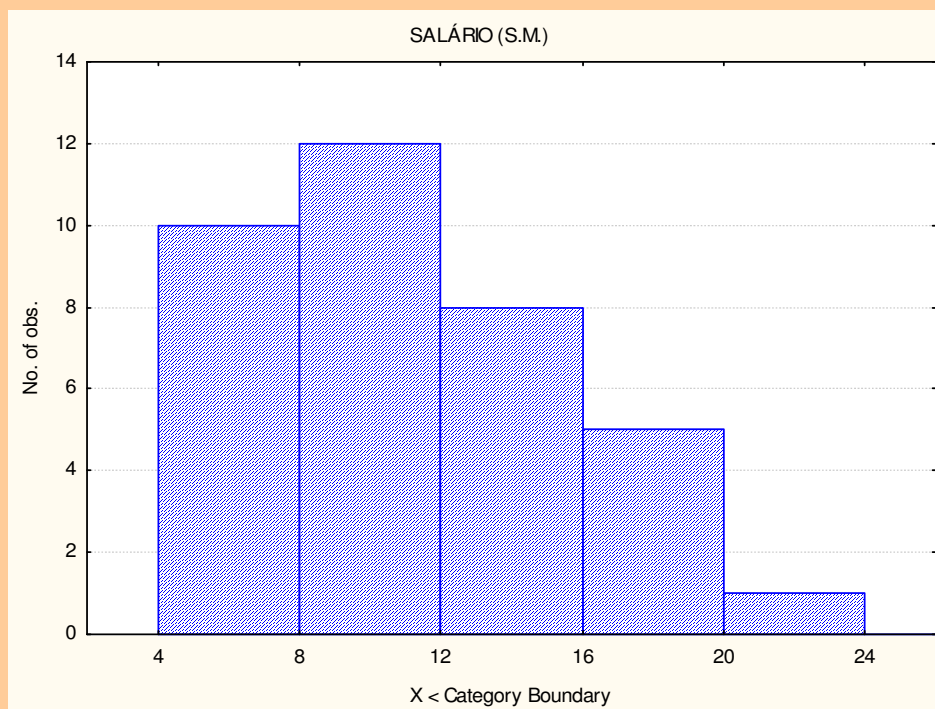


número filhos		
	Count	Percent
0	10	27,8
1	7	19,4
2	9	25,0
3	5	13,9
4	4	11,1
5	1	2,8





Faixas de salário		
	Count	Percent
4,0<=x<8,0	10	27,8
8,0<=x<12,0	12	33,3
12,0<=x<16,0	8	22,2
16,0<=x<20,0	5	13,9
20,0<=x<24,0	1	2,8



Medidas de tendência  
central  
e  
Medidas de dispersão

Os 11 alunos de uma turma de Estatística receberam as notas abaixo. Qual é a nota que mais reflete o desempenho da turma?

**7 4 2 4 5 3 6 7 4 8 9**

# Ordenação dos dados

- **antes da ordenação**

7 4 2 4 5 3 6 7 4 8 9

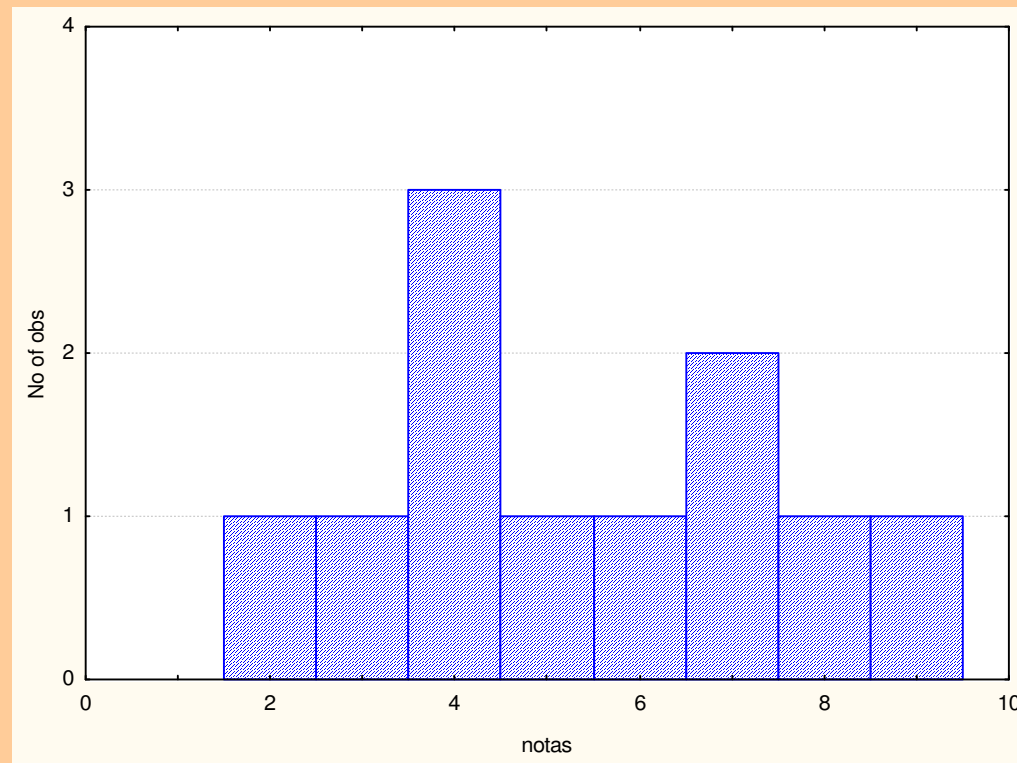
- **após a ordenação**

2 3 4 4 4 5 6 7 7 8 9

# Tabela de distribuição de frequências

notas	frequência
0	0
1	0
2	1
3	1
4	3
5	1
6	1
7	2
8	1
9	1
10	0

# Representação gráfica da tabela de distribuição de frequências



**Mas afinal, qual é a nota que  
mais reflete o desempenho  
da turma?**

# MEDIDAS DE TENDÊNCIA CENTRAL

- média
- mediana
- moda



# MÉDIA

- resultado da divisão da soma de todos os elementos de um conjunto de números pelo número total de elementos do conjunto

$$(2+3+4+4+4+5+6+7+7+8+9)/11 = 59/11$$



$$\text{média} = 5,4$$

# MÉDIA

- desvantagem: é afetada por valores extremos

Exemplo:

2      4      8      15      20

$$\text{média} = 49/5 = 9,8$$

2      4      8      15      200

$$\text{média} = 229/5 = 45,8$$

# MEDIANA

- divide uma distribuição **ordenada** de dados em 2 metades iguais
- é um índice de posição
- Posição da mediana =  $(n+1)/2$

# MEDIANA

- Quando o número de elementos (n) é ímpar a mediana corresponde ao elemento central

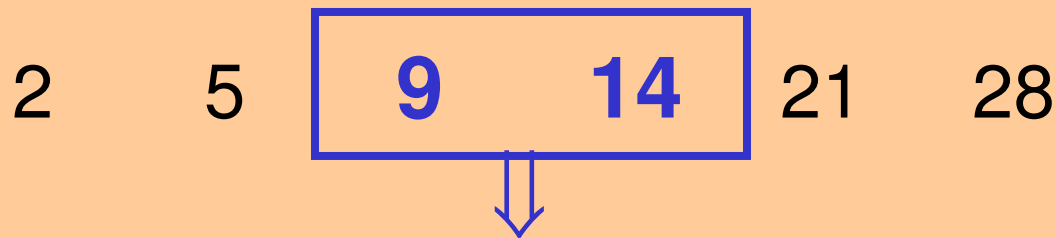
2    5    9    14    21



**Posição da mediana=  $(5+1)/2 = 3^{\circ}$  elemento**

# MEDIANA

- Quando o número de elementos é par a mediana corresponde à média aritmética dos 2 elementos centrais.



$$\text{Mediana} = (9+14)/2 = 11,5$$

$$\text{Posição da mediana} = (6+1)/2 = 3,5^\circ \text{ elemento}$$

# MEDIANA

- Turma de Estatística

2   3   4   4   4   **5**   6   7   7   8   9



**Mediana**

**Posição =  $(11+1)/2 = 6^{\circ}$  elemento**

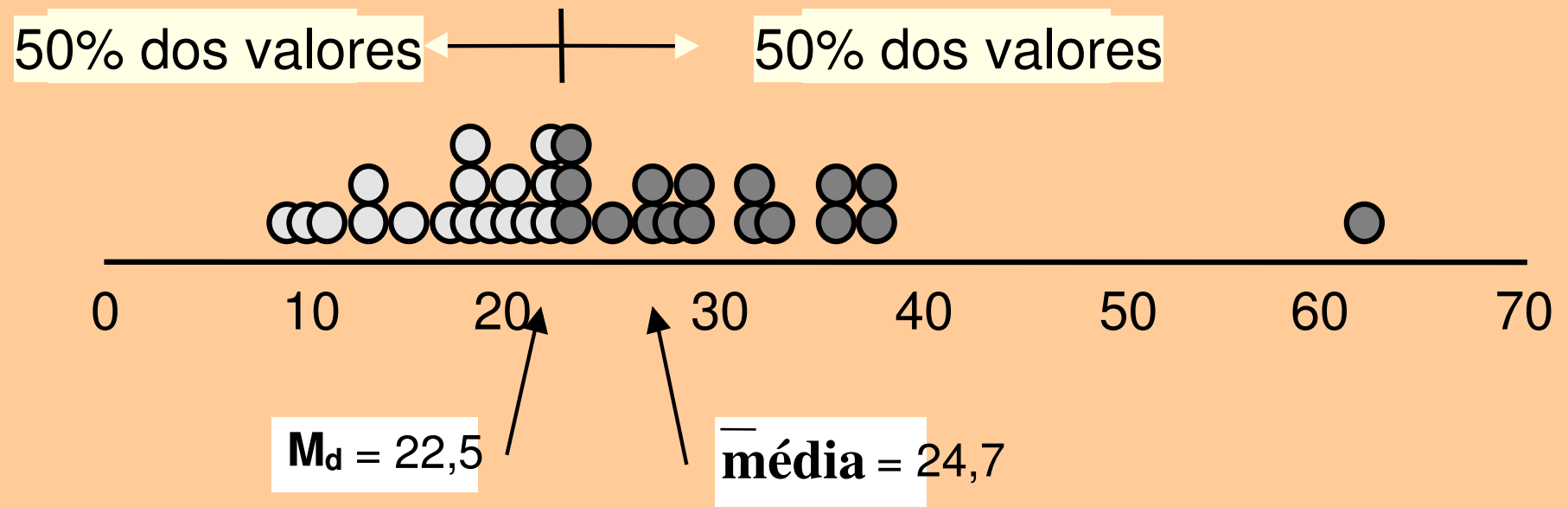
# MEDIANA

- vantagem: útil para descrever distribuições de dados com valores extremos

- Exemplo: 

2	4	8	15	20
2	4	8	15	200

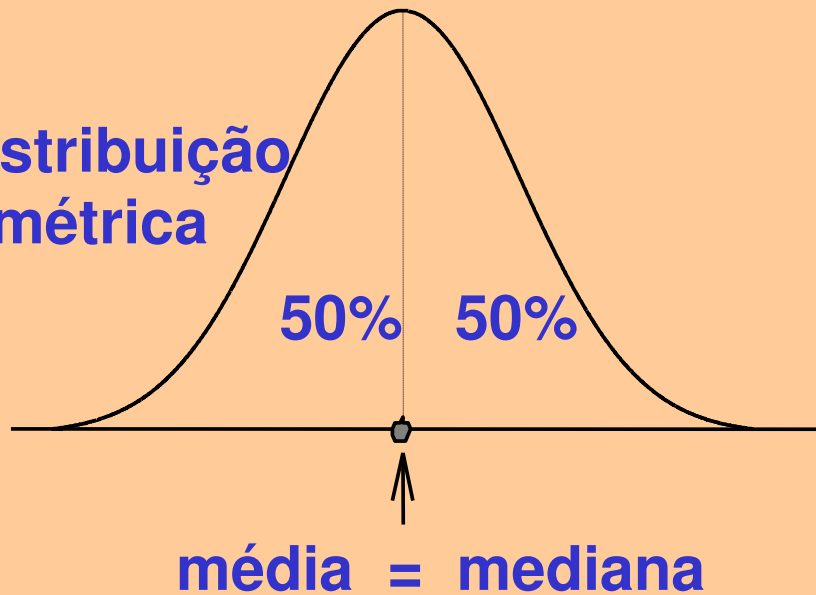
# MÉDIA E MEDIANA



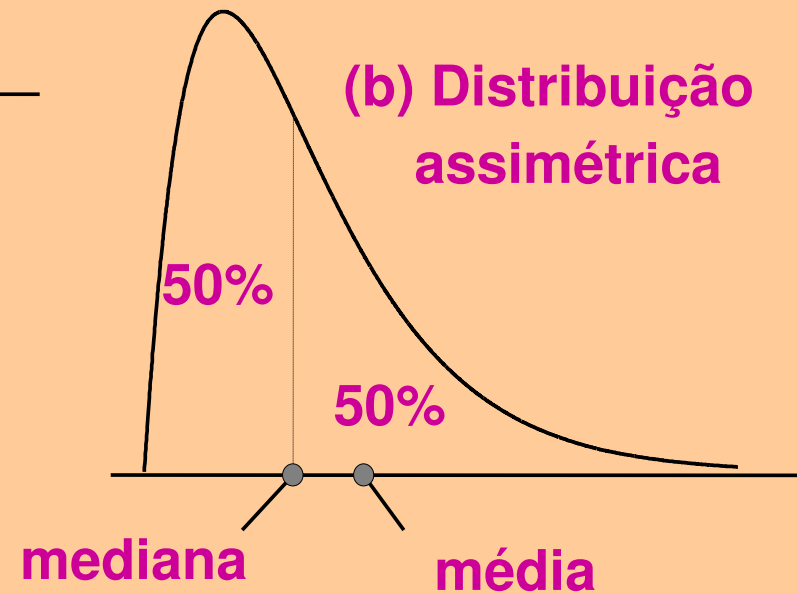


# MÉDIA E MEDIANA

(a) Distribuição  
simétrica



(b) Distribuição  
assimétrica



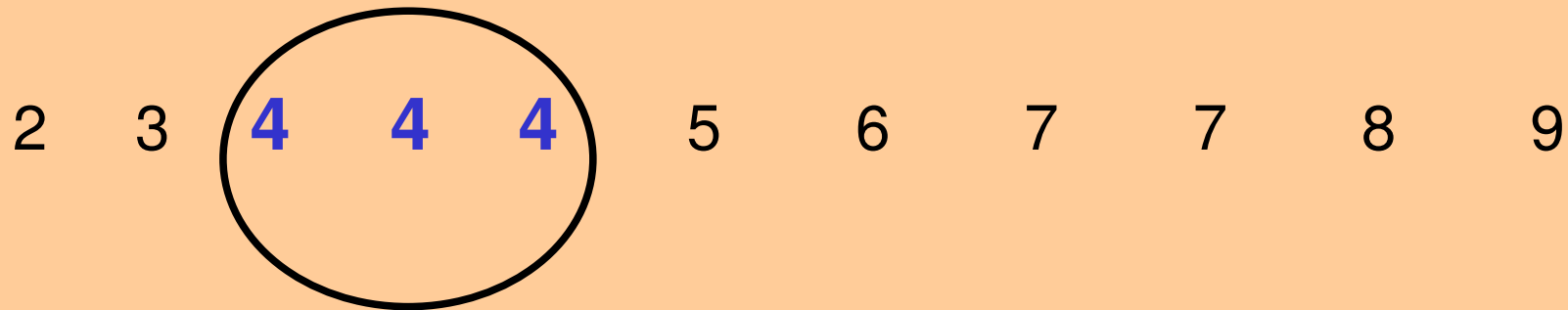
# MODA

- valor que ocorre com maior frequência
- obtida por inspeção da tabela de distribuição de frequências
- útil para medidas qualitativas

**Exemplo:** candidatos numa eleição

times de futebol num campeonato

# MODA



**Moda = 4**

# MEDIDAS SEPARATRIZES

- medidas que dividem a distribuição em partes iguais
- servem para descrever posições numa distribuição de dados

# MEDIDAS SEPARATRIZES

- PERCENTIL

divide os dados em 100 partes iguais:

$$P_1, P_2, P_3, \dots, P_{99}$$

# MEDIDAS SEPARATRIZES

- QUARTIL

divide os dados em 4 partes iguais:

$Q_1, Q_2, Q_3$

# MEDIDAS SEPARATRIZES

- percentil 50 = mediana = segundo quartil ( $Q_2$ )
- percentil 25 = primeiro quartil ( $Q_1$ )
- percentil 75 = terceiro quartil ( $Q_3$ )

# MEDIDAS SEPARATRIZES

Posição do  $Q_1 = \frac{1}{4} (n+1)$

Posição do  $Q_2 = \frac{1}{2} (n+1)$

Posição do  $Q_3 = \frac{3}{4} (n+1)$



# Cálculo dos quartis

0,9 1,0 1,7 2,9 3,1 5,3 5,5 12,2 12,9 14,0 33,6 → n=11

**Posição do  $Q_1$**  =  $\frac{1}{4} (11+1) = 3^{\circ}$  elemento  
→  $Q_1 = 1,7$

**Posição do  $Q_2$**  =  $\frac{1}{2} (11+1) = 6^{\circ}$  elemento  
→  $Q_2 = 5,3$

**Posição do  $Q_3$**  =  $\frac{3}{4} (11+1) = 9^{\circ}$  elemento  
→  $Q_3 = 12,9$

# Cálculo dos quartis

19 20 21 25 30 31 33 37 61 77

→ n=10

**Posição do  $Q_1$**  =  $\frac{1}{4} (10+1) = 2,75^\circ$  elemento

→  $Q_1 = (20+21)/2 = 20,5$

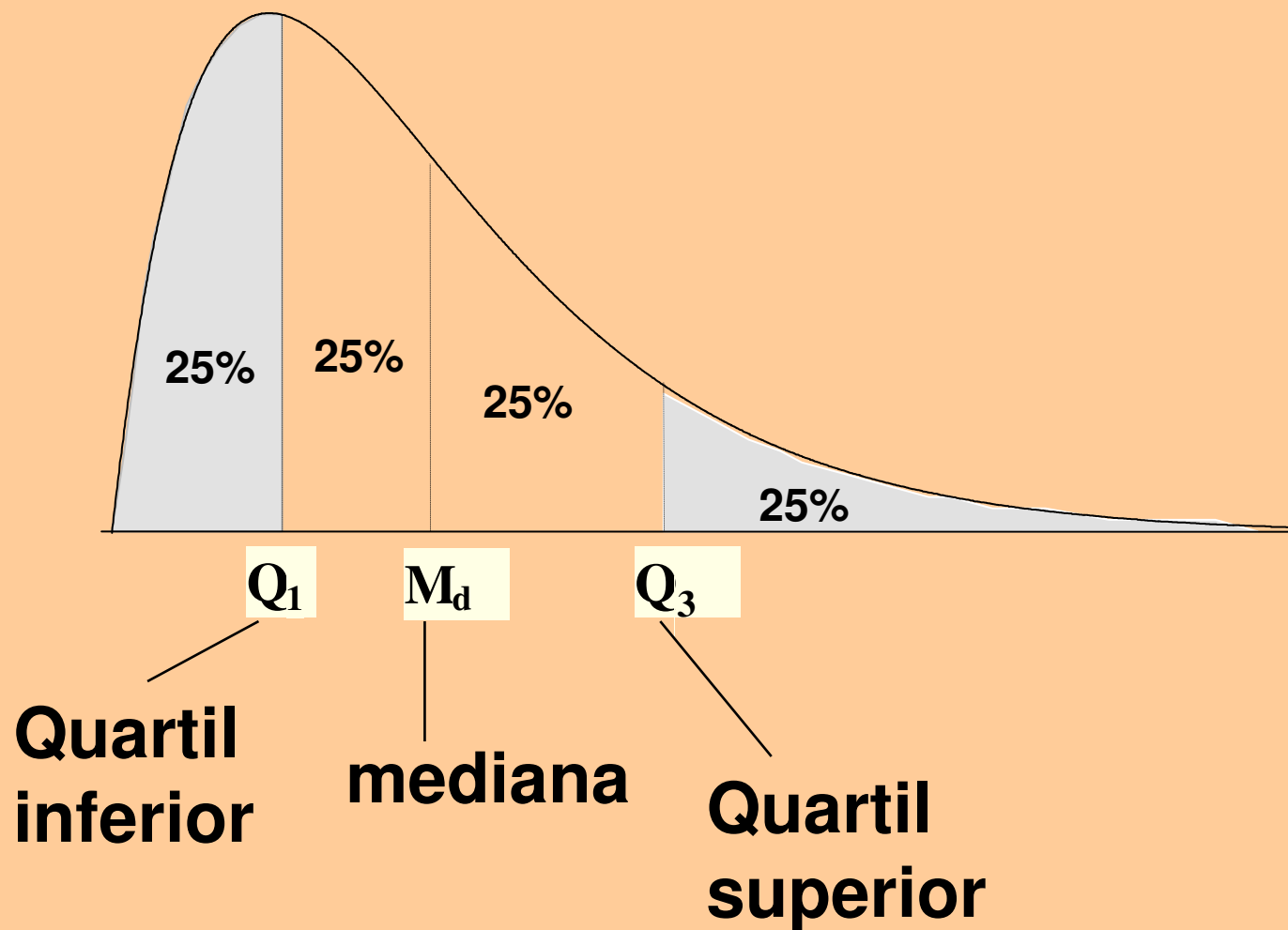
**Posição do  $Q_2$**  =  $\frac{1}{2} (10+1) = 5,5^\circ$  elemento

→  $Q_2 = (30+31)/2 = 30,5$

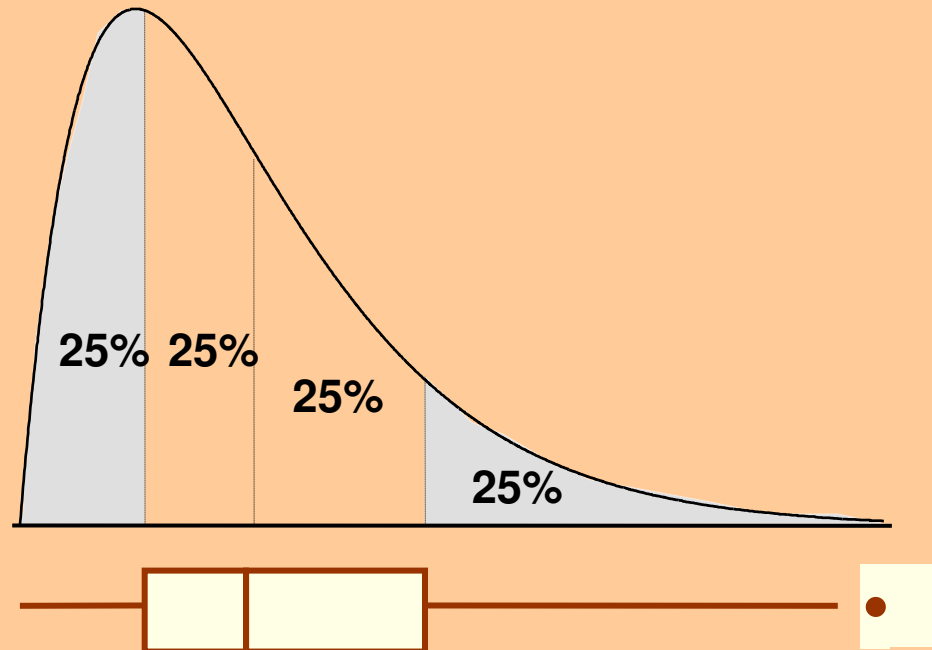
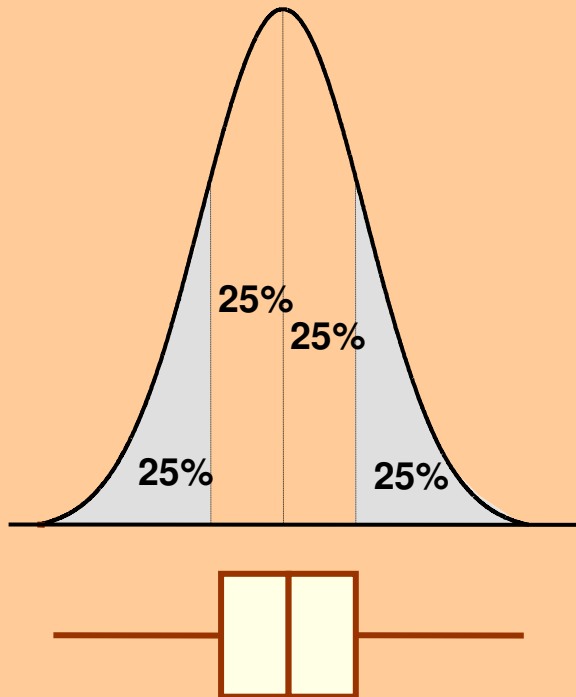
**Posição do  $Q_3$**  =  $\frac{3}{4} (10+1) = 8,25^\circ$  elemento

→  $Q_3 = (37+61)/2 = 49$

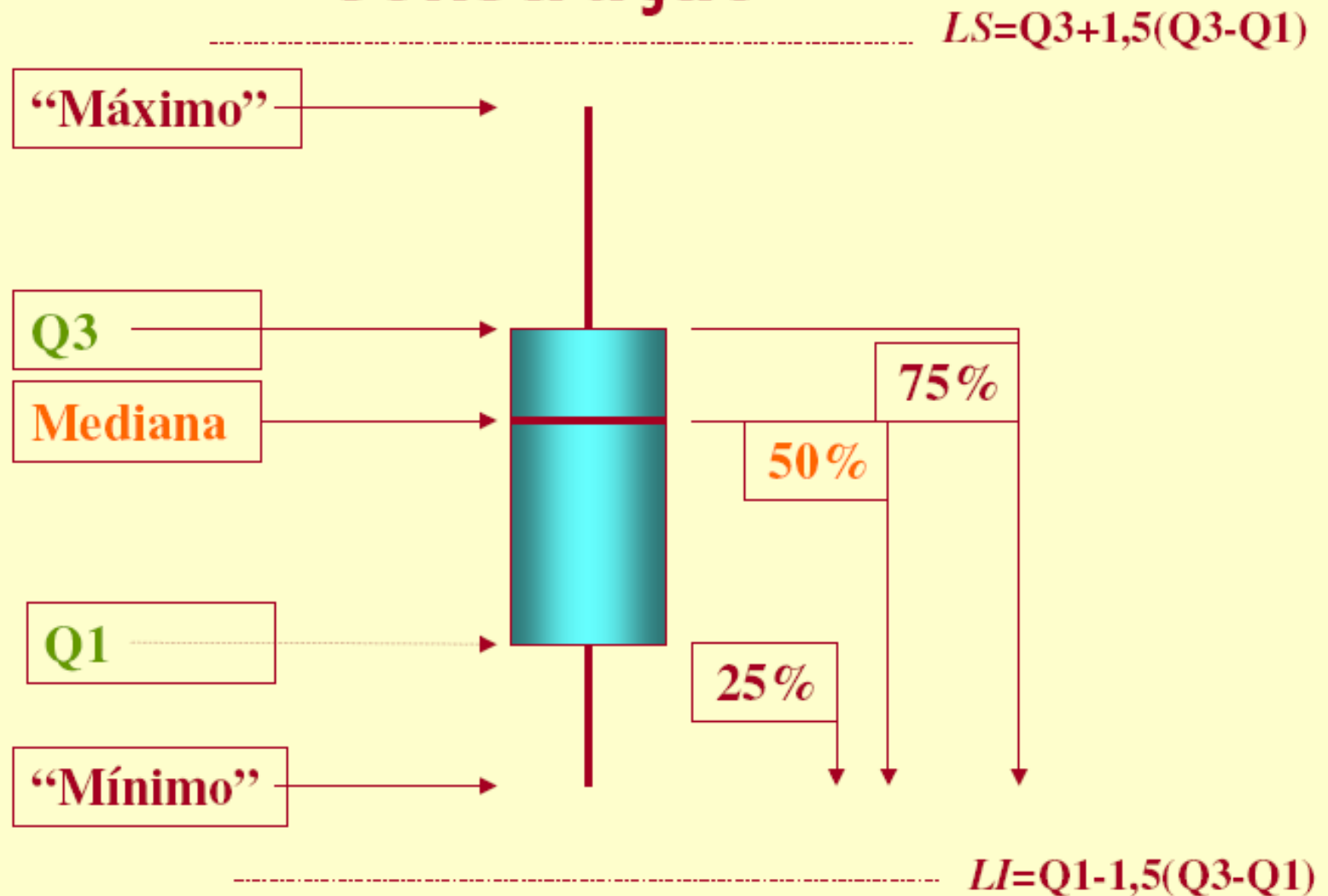
# QUARTIS NUMA DISTRIBUIÇÃO



# BOX-PLOT



# Construção



“Máximo” é o maior valor menor que  $LS$ ;

“Mínimo” é o menor valor maior que  $LI$ .

# Exemplo: Tempo de sobrevivência (dias)

Dados ordenados (n=36)

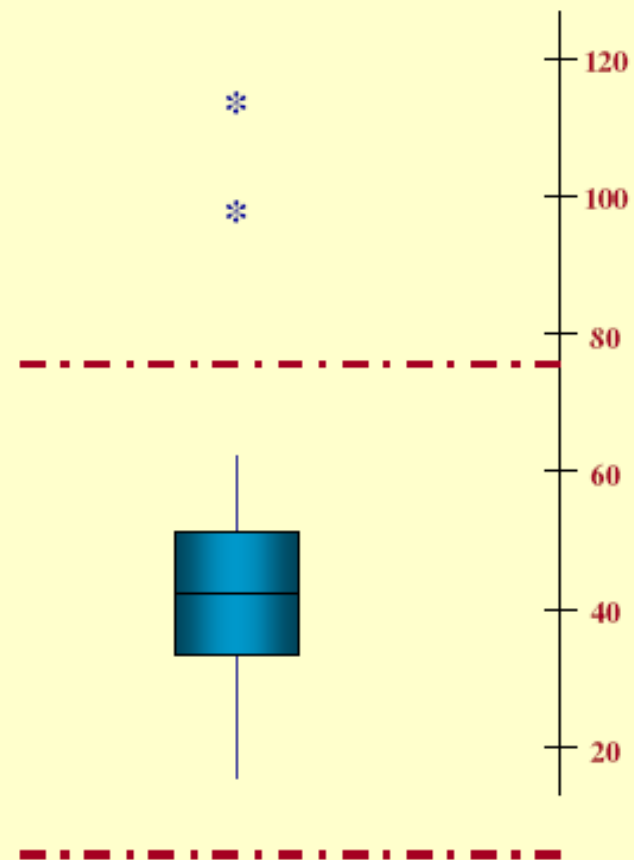
18	21	21	23	23	25
27	29	30	31	32	32
32	34	35	36	38	41
42	42	43	44	45	46
46	47	48	50	54	56
57	58	60	61	98	116

md = 41,5    Q1 = 30,25    Q3 = 49,5

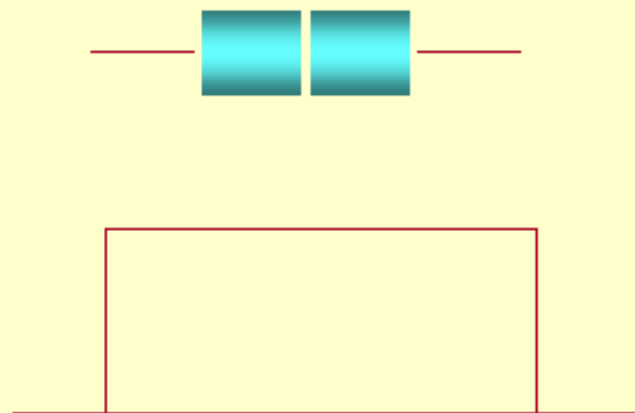
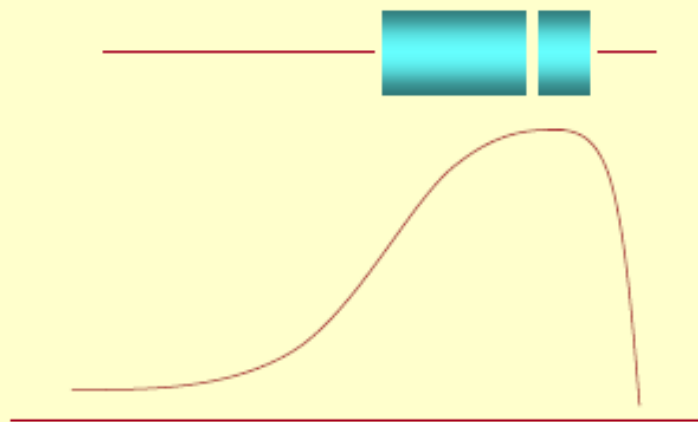
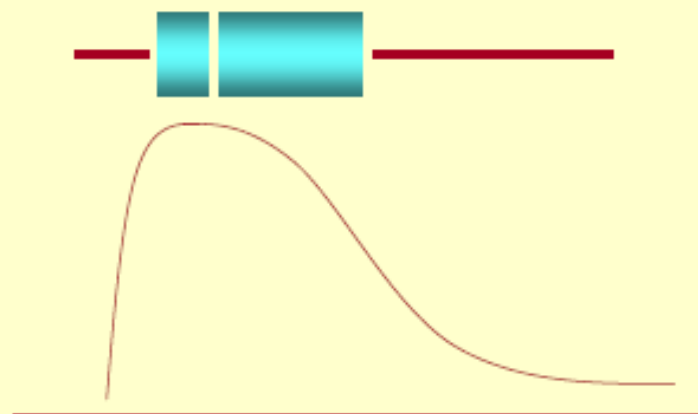
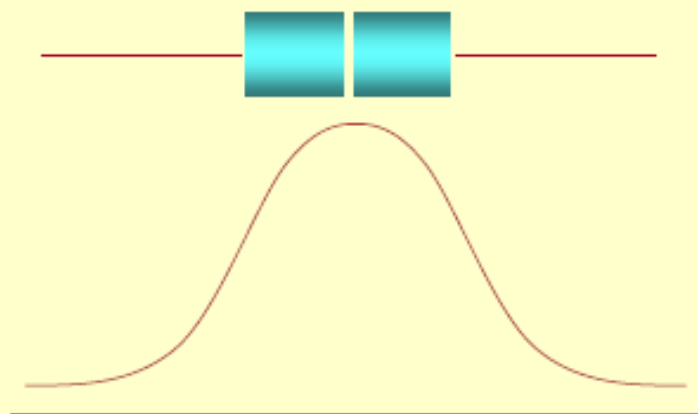
*Observações discrepantes?*

$$LI = Q1 - 1,5(Q3 - Q1) = 1,38$$

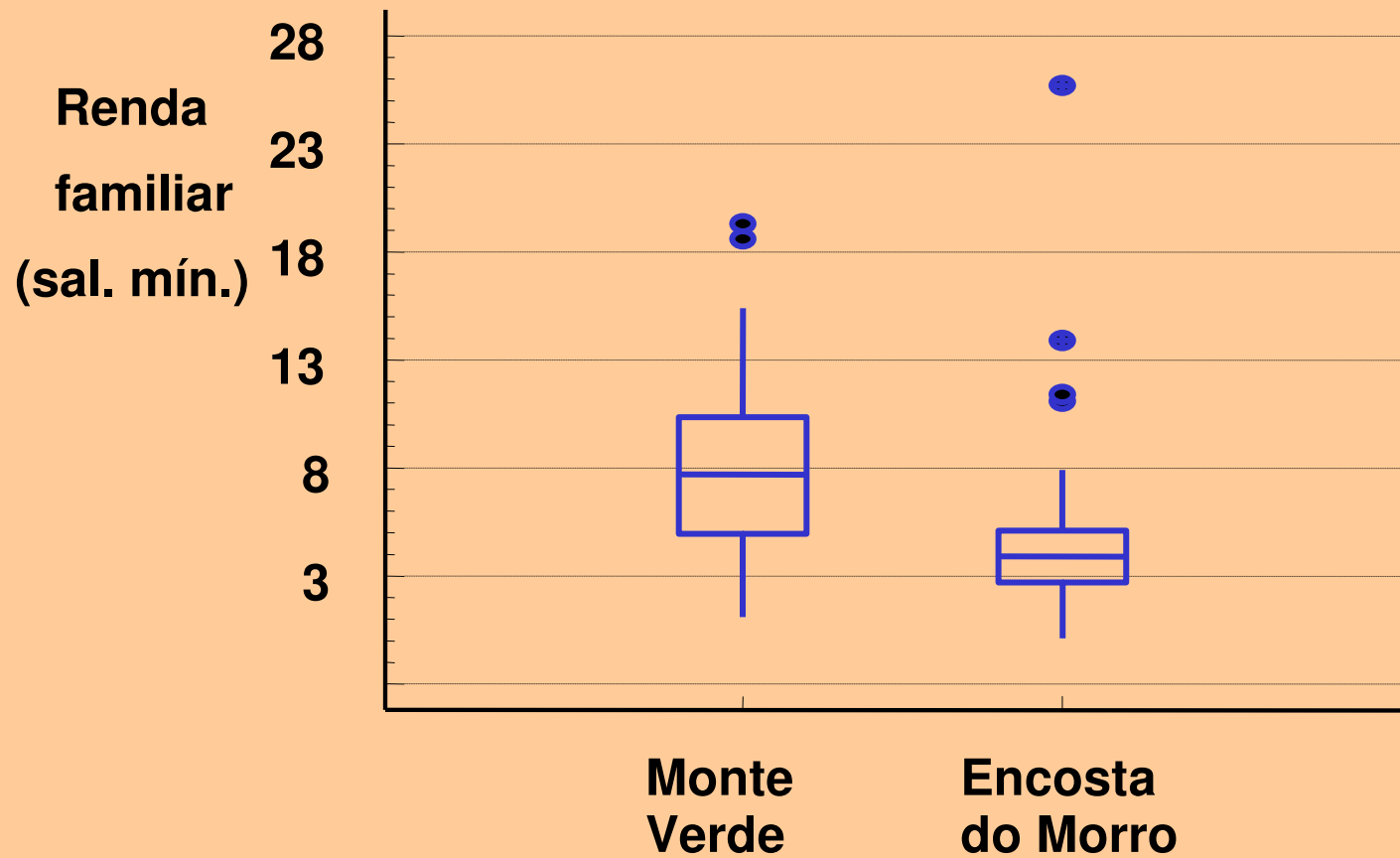
$$LS = Q3 + 1,5(Q3 - Q1) = 78,38$$



# Forma da Distribuição

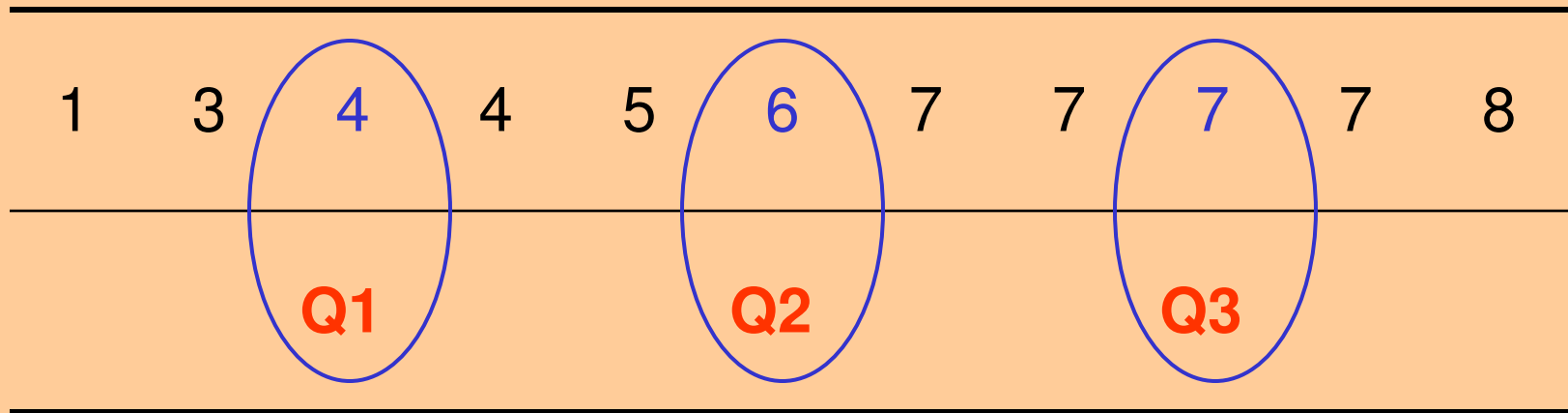


# BOX-PLOT

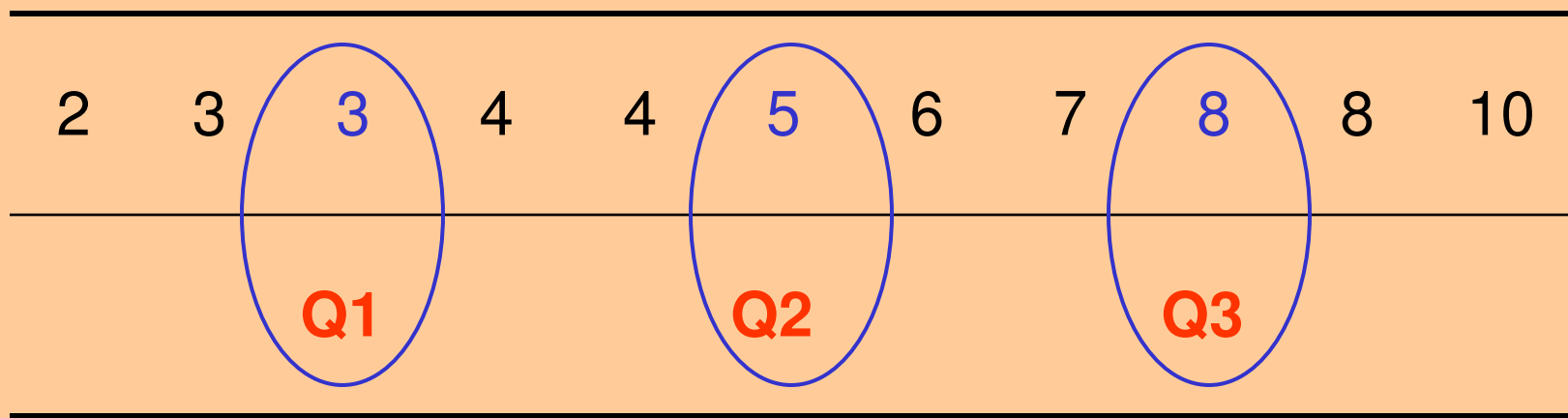




## Notas turma A



## Notas turma B



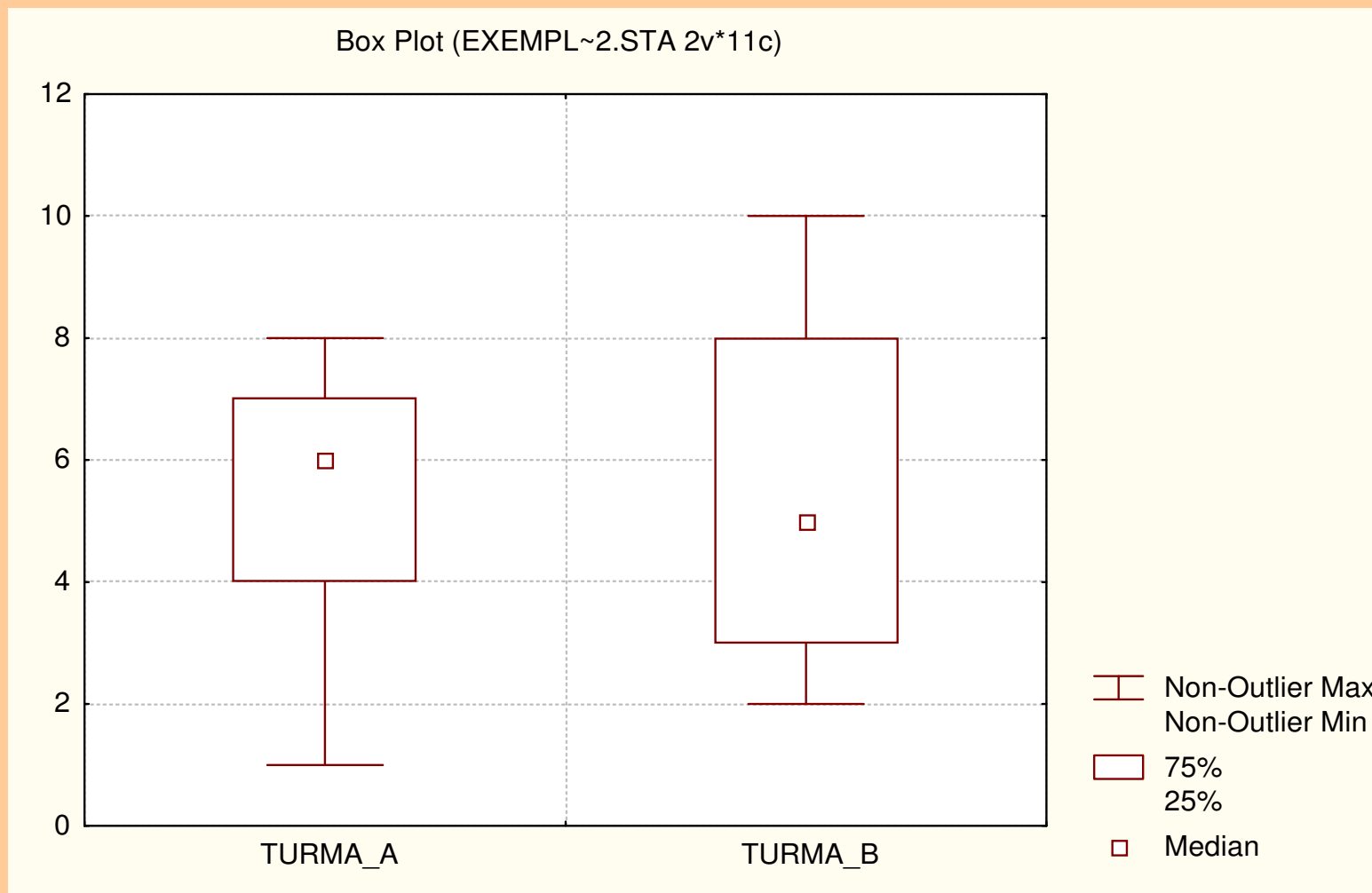


Figura 1 – Nota das turmas A e B expressa através de box-plot.

# MEDIDAS DE DISPERSÃO

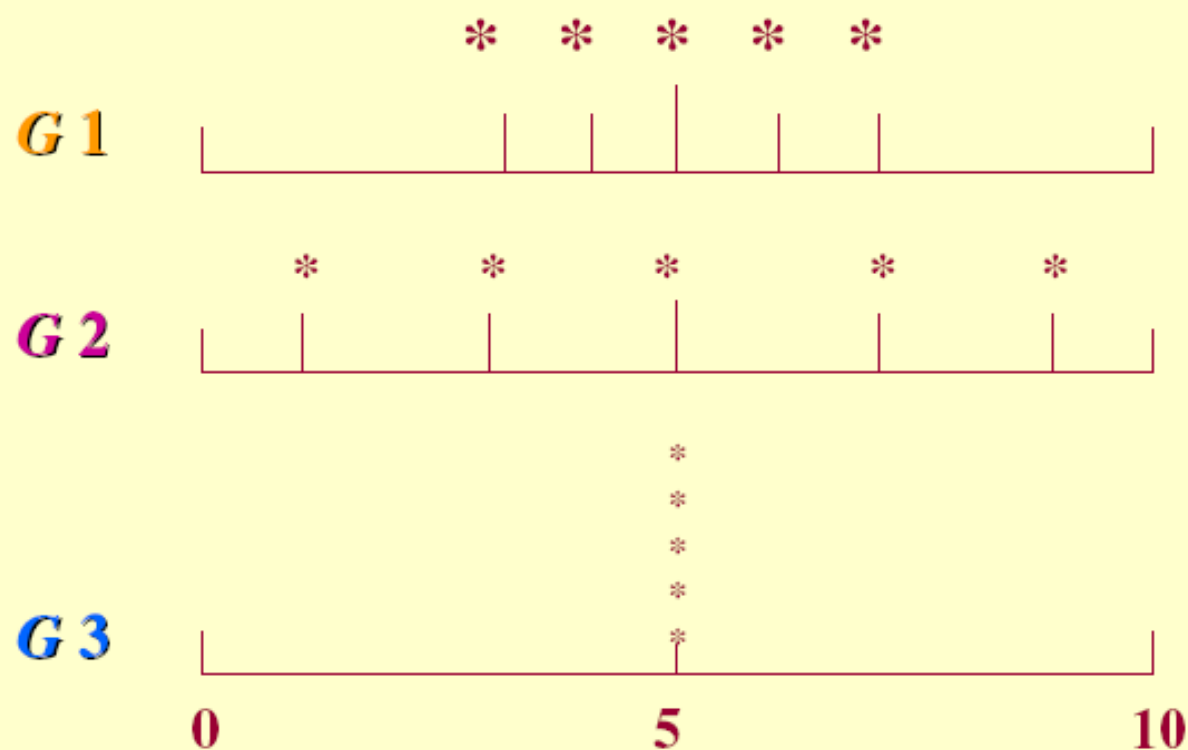
- amplitude total ou intervalo
- Intervalo interquartil
- variância
- desvio-padrão

**Exemplo 2:** Considere as notas de um teste de 3 grupos de alunos

**Grupo 1:** 3,4,5,6,7

**Grupo 2:** 1, 3, 5, 7, 9

**Grupo 3:** 5,5,5,5,5



Temos:  $\bar{x}_1 = \bar{x}_2 = \bar{x}_3 = 5$  e  $md_1 = md_2 = md_3 = 5$

# Medidas de Dispersão

**Finalidade:** encontrar um valor que resuma a variabilidade de um conjunto de dados

- **Amplitude (A):**

$$A = \textit{máx} - \textit{min}$$

Para os grupos anteriores, temos:

Grupo 1,  $A = 4$

Grupo 2,  $A = 8$

Grupo 3,  $A = 0$

# Intervalo interquartil

É a diferença entre Q3 e Q1

$$IQ = Q3 - Q1$$

19 20 21 25 30 31 33 37 61 77

→ n=10

$$Q_1 = 20,5 \quad Q_2 = 30,5 \quad Q_3 = 49$$

$$IQ = 49 - 20,5 = 28,5$$

# DESVIO-PADRÃO

- é a medida de dispersão mais usada
- determina a dispersão de todos os valores em relação à média
- as distâncias em relação à média são elevadas ao quadrado

- Variância:

$$\text{Variância} = s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n-1} = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1}$$

- Desvio padrão:

$$\text{Desvio Padrão} = s = \sqrt{\text{Variância}}$$



## Cálculo para os grupos:

$$\mathbf{G1: } s^2 = \frac{(3-5)^2 + (4-5)^2 + (5-5)^2 + (6-5)^2 + (7-5)^2}{4}$$

$$\Rightarrow s^2 = 10/4 = 2,5 \Rightarrow s = 1,58$$

$$\mathbf{G2: } s^2 = 10 \Rightarrow s = 3,16$$

$$\mathbf{G3: } s^2 = 0 \Rightarrow s = 0$$

## • Coeficiente de Variação (*CV*)

- é uma medida de dispersão relativa
- elimina o efeito da magnitude dos dados
- exprime a variabilidade em relação à média

$$CV = \frac{s}{\bar{x}} \times 100 \%$$

## Exemplo 3:

Altura e peso de alunos

	Média	Desvio Padrão	Coef. de Variação
Altura	1,143m	0,063m	5,5%
Peso	50 kg	6kg	12%

**Conclusão:** Os alunos são, aproximadamente, duas vezes mais dispersos quanto ao peso do que quanto à altura.

## Exemplo 4:

Altura (em *cm*) de uma amostra de recém-nascidos e de uma amostra de adolescentes

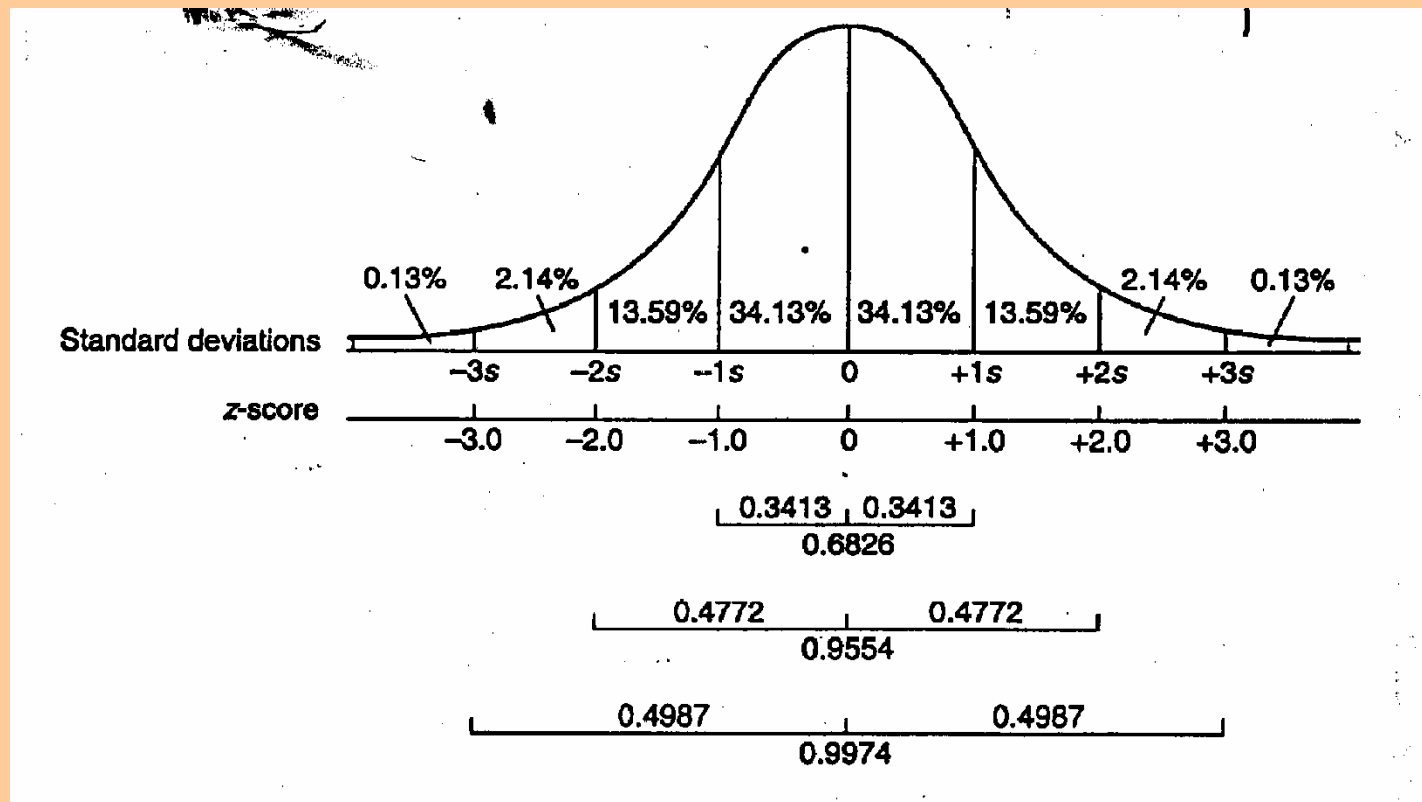
	Média	Desvio padrão	Coef. de variação
Recém-nascidos	50	6	12%
Adolescentes	160	16	10%

**Conclusão:** Em relação às médias, as alturas dos adolescentes e dos recém-nascidos apresentam variabilidade quase iguais.

# A DISTRIBUIÇÃO NORMAL

- muitos fenômenos biológicos, psicológicos e sociais seguem uma distribuição normal

# CURVA DE GAUSS\_(Normal)



# Faixa de normalidade

- média aritmética  $\pm$  desvio-padrão
- corresponde à aproximadamente 68% dos indivíduos da amostra

## Exemplo: pesos obtidos em 2 grupos de atletas

	<i>Grupo A</i>	<i>Grupo B</i>
	78	65
	80	69
	82	78
	85	85
	85	85
	85	93
	86	96
	88	98
<i>Soma</i>	669	669
<i>Média</i>	83,6	83,6
<i>Mediana</i>	85	85
<i>Moda</i>	85	85
<i>N</i>	8	8



# DESVIO-PADRÃO

peso grupo A	d = peso - média	d <sup>2</sup>
78	78 - 83,6= -5,6	(-5,6) <sup>2</sup> = 31,36
80	80 - 83,6=-3,6	(-3,6) <sup>2</sup> = 12,96
82	82 - 83,6=-1,6	(-1,6) <sup>2</sup> = 2,56
85	85 - 83,6= 1,4	(1,4) <sup>2</sup> = 1,96
85	85 - 83,6= 1,4	(1,4) <sup>2</sup> = 1,96
85	85 - 83,6= 1,4	(1,4) <sup>2</sup> = 1,96
86	86 - 83,6= 2,4	(2,4) <sup>2</sup> = 5,76
88	88 - 83,6= 4,4	(4,4) <sup>2</sup> = 19,36
		<b>soma = 77,88</b>

**Variância = (soma d<sup>2</sup>/ (n-1)) = 77,88/(8-1) = 11,1 kg<sup>2</sup>**

**Desvio-padrão = raiz (variância) = raiz (11,1) = 3,3 Kg**

# DESVIO-PADRÃO

peso grupo B	d = peso - média	d <sup>2</sup>
65	65 - 83,6 = -18,6	(-18,6) <sup>2</sup> = 31,36
69	69 - 83,6 = -14,6	(-14,6) <sup>2</sup> = 12,96
78	78 - 83,6 = -5,6	(- 5,6) <sup>2</sup> = 2,56
85	85 - 83,6 = 1,4	(1,4) <sup>2</sup> = 1,96
85	85 - 83,6 = 1,4	(1,4) <sup>2</sup> = 1,96
93	93 - 83,6 = 9,4	(9,4) <sup>2</sup> = 1,96
96	96 - 83,6 = 12,4	(12,4) <sup>2</sup> = 5,76
98	98 - 83,6 = 14,4	(14,4) <sup>2</sup> = 19,36
		soma = 1043,88

**Variância = (soma d<sup>2</sup>/ (n-1)) = 1043,88/(8-1) = 149,12 kg<sup>2</sup>**

**Desvio-padrão = raiz (variância) = raiz (149,12) = 12,2 Kg**

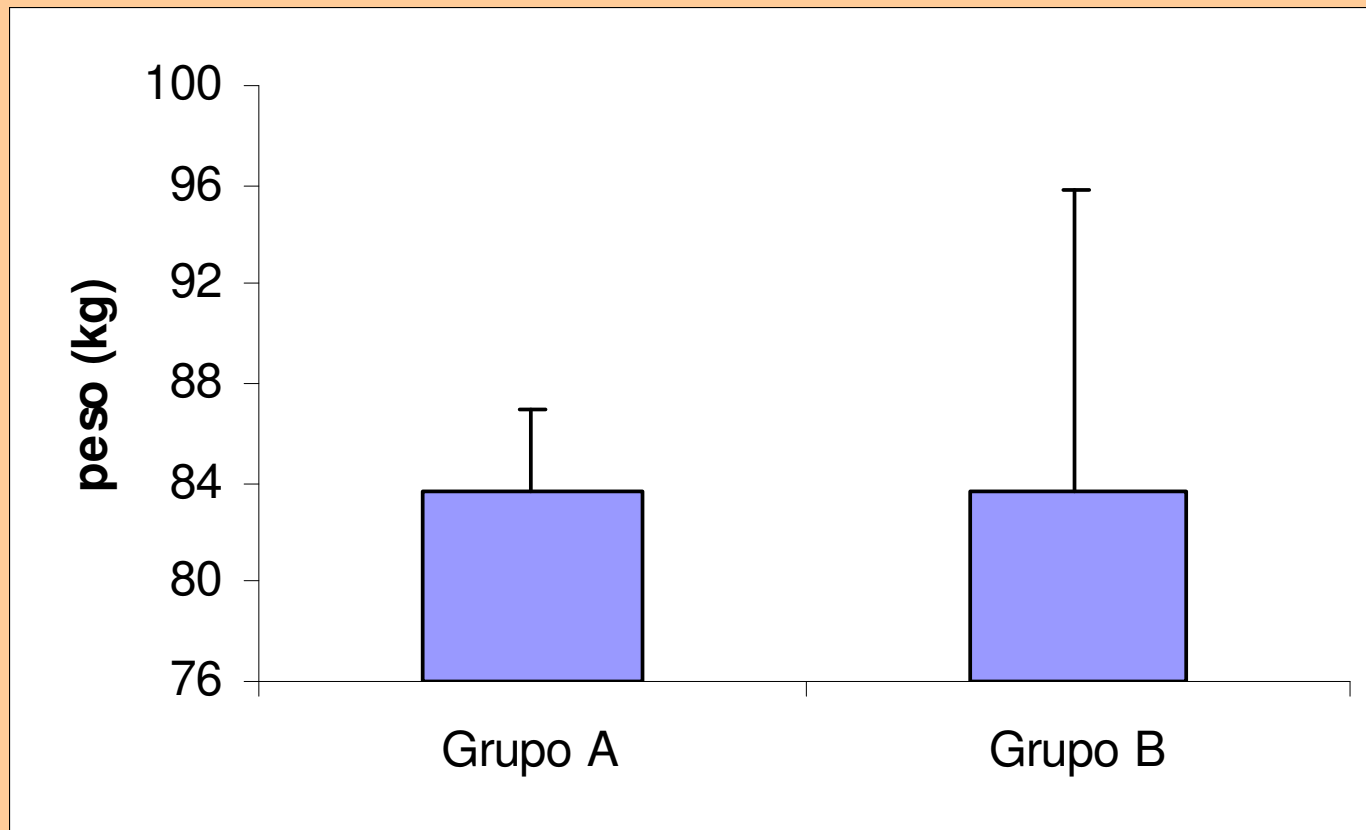
# **Faixa de normalidade**

## **Grupo A**

- Média = 83,6Kg; desvio-padrão= 3,3 Kg
- Faixa de normalidade=  $83,6 \pm 3,3$  Kg

## **Grupo B**

- Média = 83,6Kg; desvio-padrão= 12,2 Kg  
Faixa de normalidade=  $83,6 \pm 12,2$  Kg



**Figura 1 – Peso dos grupos A e B.  
Dados expressos em média  $\pm$  dp.**