

3 – Estatística p/ Descrição, Exploração e Comparação de Dados (Triola 10^a ed.)

C V D O T

C entro	Valor representativo ou médio
V ariação	Quanto os valores variam entre eles
D istribuição	Natureza dos dados
O utliers	Valores que se localizam muito longe
T empo	Dados que mudam com o tempo

Capítulo 3 – Triola, 10^a. Ed.

Conteúdo

3.1 Visão Geral

3.2 Medidas de Centro (resumo)

3.3 Medidas de Variação

3.4 Medidas de Posição relativa

3.5 Análise Exploratória de Dados (AED)

Continuação do Problema do Capítulo 2

A premiação do “Oscar” envolve discriminação com base na idade?

Tabela 2-1 Prêmios da Academia: Idade das Melhores Atrizes e dos Melhores Atores

As idades (em anos) estão listadas em ordem, começando com a primeira cerimônia de premiação.

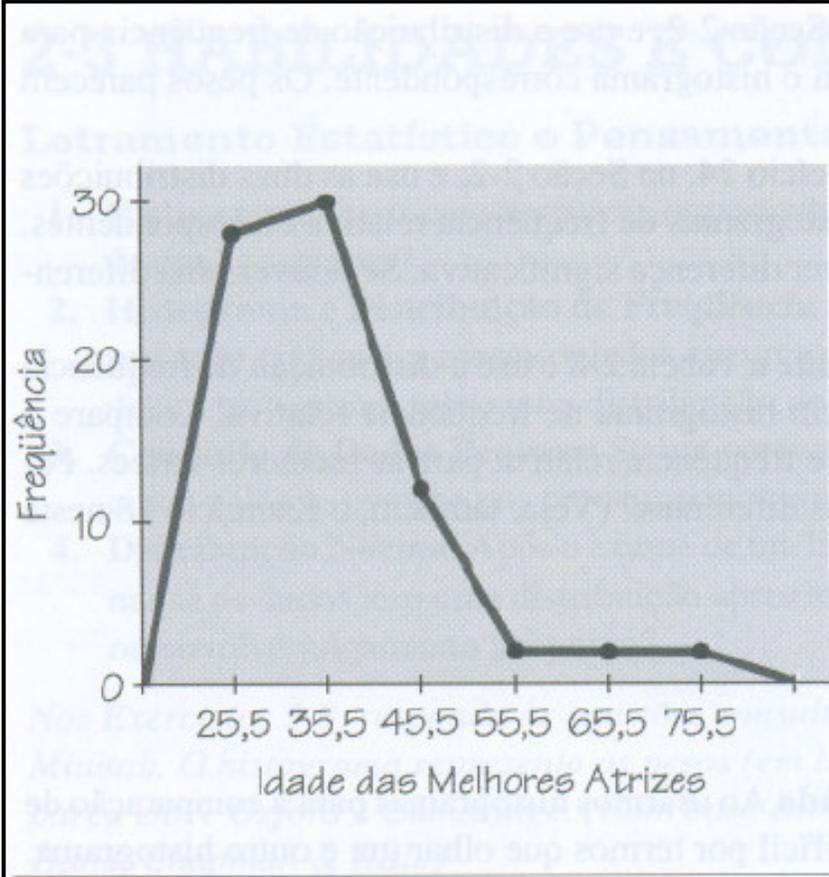


FIGURA 2-4 Polígono de Frequências

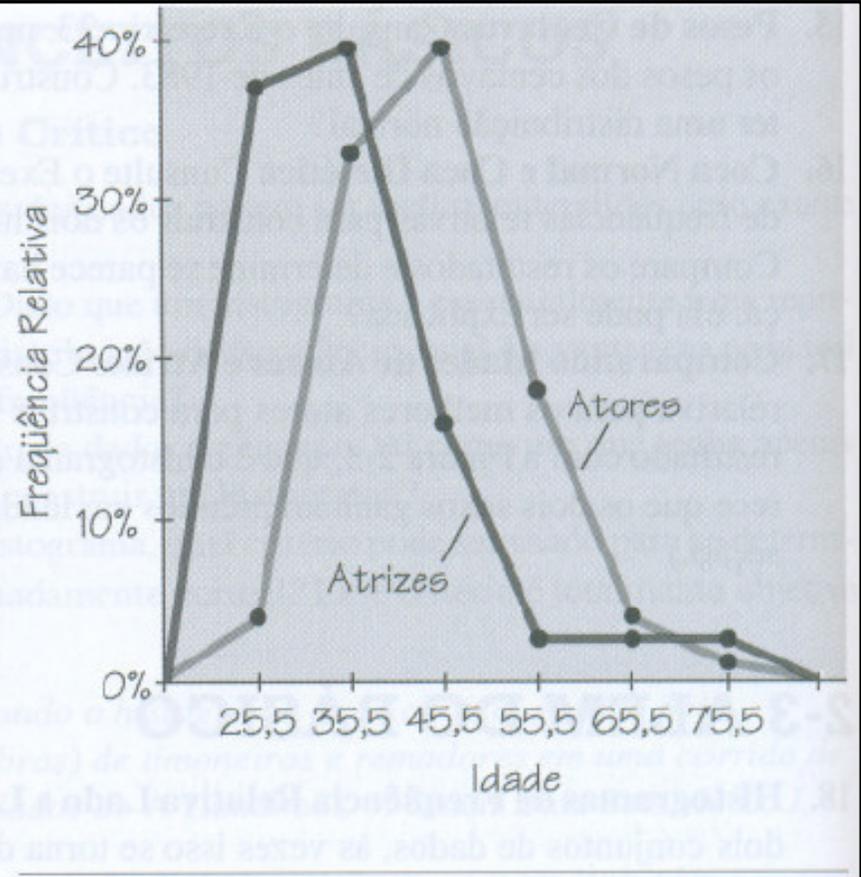


FIGURA 2-5 Polígonos de Frequências Relativas

Capítulo 3 – Triola, 10^a. Ed.

Conteúdo

3.1 Visão Geral

Princípio do uso moderno da estatística

(Triola, p62):

Não é importante memorizar fórmulas. Em vez disto devemos nos concentrar na obtenção de resultados com o uso da tecnologia e obter o sentido prático destes resultados através do **pensamento crítico**.

Capítulo 3 – Triola, 10^a. Ed.

Conteúdo

3.1 Visão Geral

3.2 Medidas de Centro (resumo)

3.3 Medidas de Variação

3.4 Medidas de Posição relativa

3.5 Análise Exploratória de Dados (AED)

3.2 Medidas de Centro (Lembre-se C.V.D.O.T.)

É um valor no centro (ou no interior) do conjunto de dados.

Média (aritmética)

Σ	Indica adição de um conjunto de valores.
x	É a variável, representa os valores individuais dos dados.
n	Número de valores em uma <i>amostra</i> .
N	Número de valores em uma <i>população</i> .
$\bar{x} = \frac{\Sigma x}{n}$	\bar{x} barra: Média de um conjunto de valores <i>amostrais</i> .
$\mu = \frac{\Sigma x}{N}$	μ mi: Média de um conjunto de valores em uma <i>população</i> .
\tilde{x}	\tilde{x} til: Mediana (definição a seguir).

3.2 Medidas de Centro

Mediana \tilde{x} (x til)

É o **valor** do meio quando os dados são arranjados em ordem crescente (ou decrescente) em magnitude.

1. Se o número de valores (n) for ímpar, a mediana é o número **localizado** no meio exato dos dados
2. Se o número de valores for par, a mediana é dada pela média dos dois números **localizados** no meio.

3.2 Medidas de Centro

Moda representada por M

É o valor que ocorre mais frequentemente.

Quando dois valores ocorrem com mesma freq., cada um é uma moda e o conjunto de dados é **bimodal**.

Quando mais de dois valores ocorrem com a mesma freq. cada um é uma moda, o conjunto é **multimodal**.

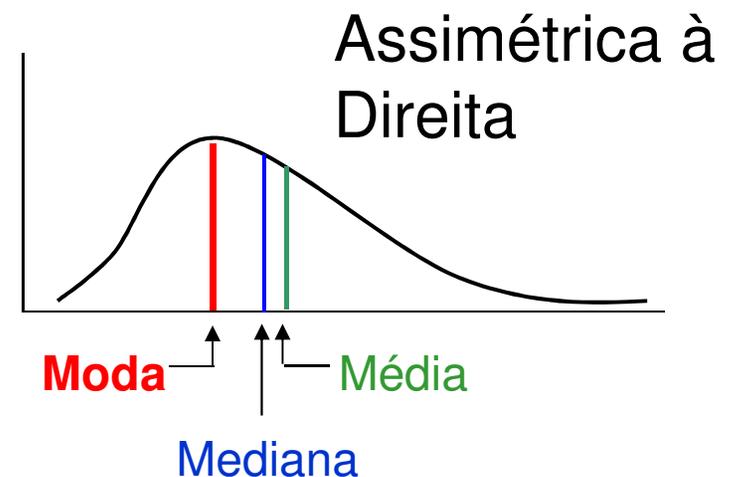
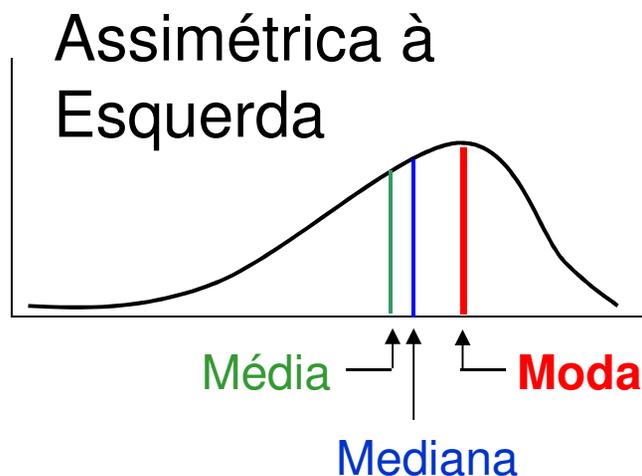
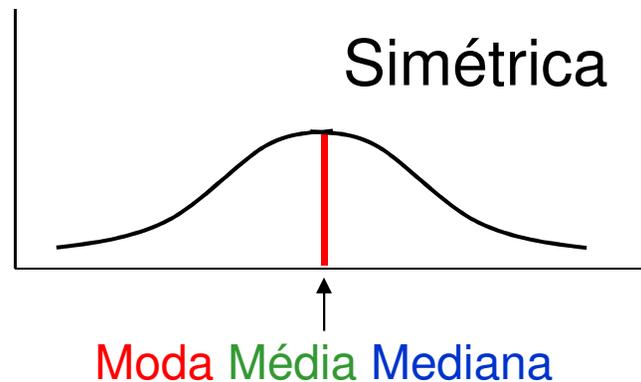
3.2 Medidas de Centro

Ponto médio raramente utilizado

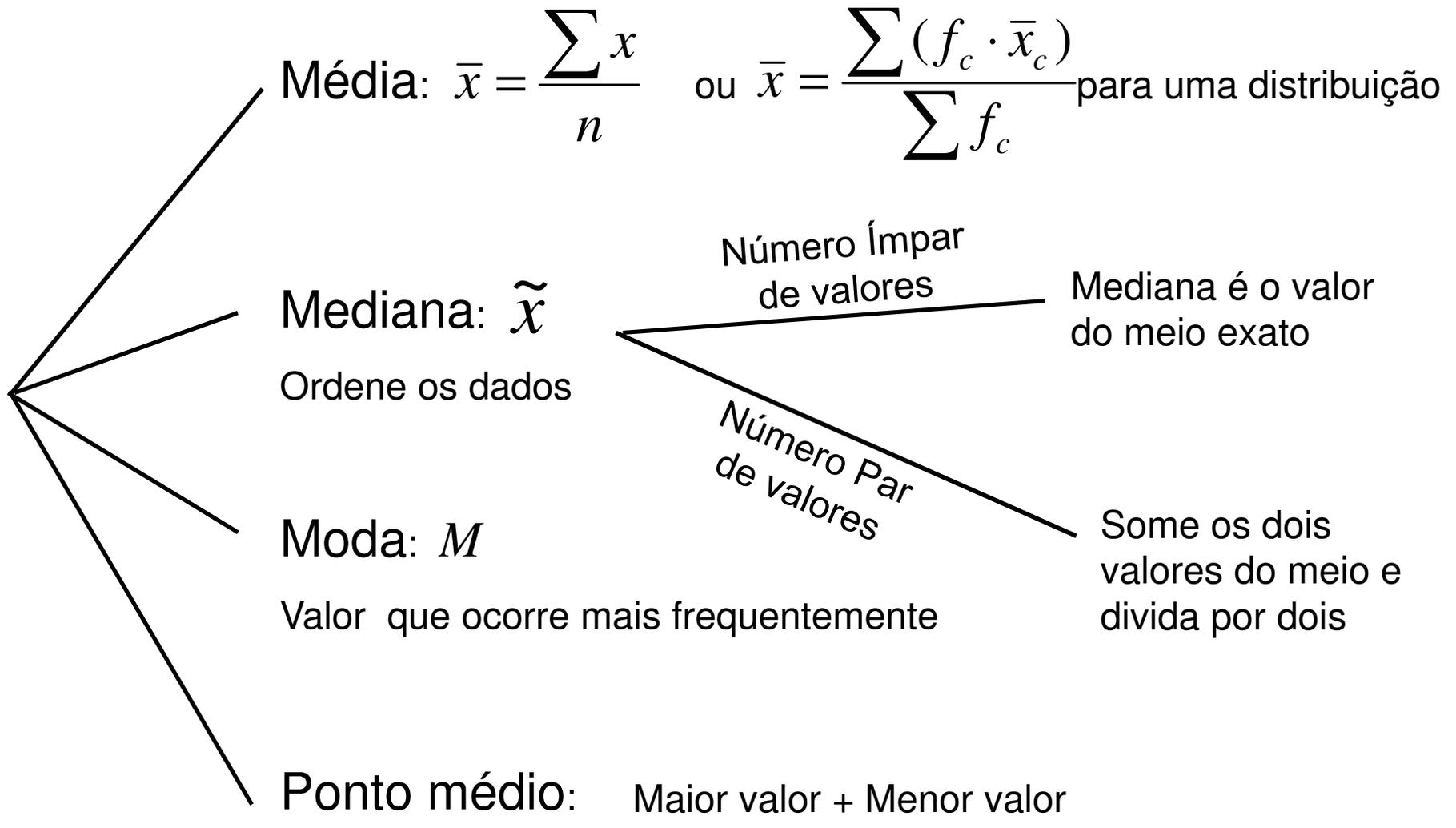
Ponto médio = $(\text{maior valor} + \text{menor valor})/2$

Uma comparação da média, mediana e moda pode revelar informação sobre a característica da simetria.

Assimetria: A distribuição de dados é assimétrica quando se estende mais para um lado do que para o outro.



Para encontrar medidas de centro (Resumo)



Capítulo 3 – Triola, 10^a. Ed.

Conteúdo

3.1 Visão Geral

3.2 Medidas de Centro (resumo)

3.3 Medidas de Variação

3.4 Medidas de Posição relativa

3.5 Análise Exploratória de Dados (AED)

3.3 Medidas de variação - CVDOT

Um dos assuntos mais importantes (Medidas de Risco)

Leia a Seção

Aprenda a calcular

Entenda o raciocínio

Amplitude: $M_v - m_v$

Desvio padrão de uma amostra: $s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$

Desvio padrão de uma população: $\sigma = \sqrt{\frac{\sum (x - \mu)^2}{N}}$

Variância: amostral s^2

populacional σ^2

Coeficiente de variação: $CV = \frac{s}{\bar{x}} 100\%$ $CV = \frac{\sigma}{\mu} 100\%$

3.3 Medidas de variação – CVDOT

Regras empíricas

Para ter uma idéia de um desvio padrão (pág. 80):

$$s = \frac{\textit{amplitude}}{4}$$

Para interpretar um valor conhecido do desvio padrão:

Valor mínimo “usual” \approx (média) $- 2 \times$ (desvio padrão)

Valor máximo “usual” \approx (média) $+ 2 \times$ (desvio padrão)

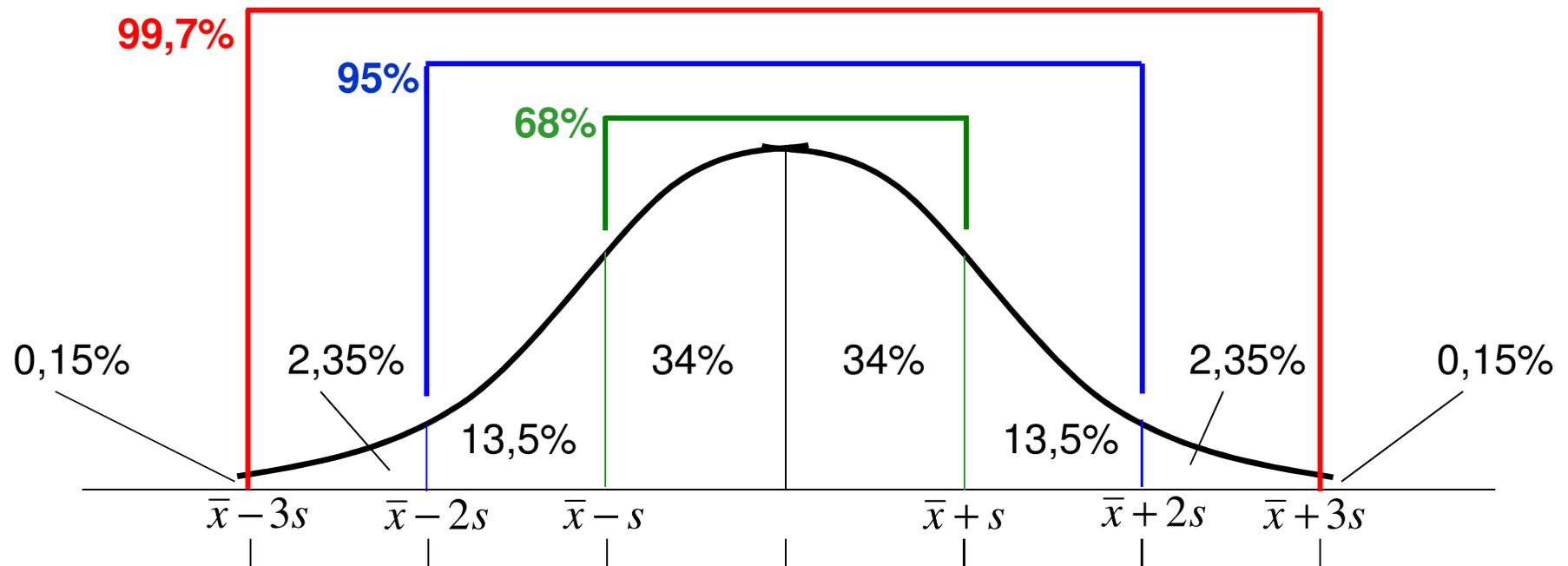
3.3 Medidas de variação – CVDOT

Regra empírica 68 – 95 – 99,7 para dados com uma distribuição em forma de sino (pág 81):

Cerca de **68 %** de todos os valores ficam a 1 desvio padrão da média

Cerca de **95 %** de todos os valores ficam a 2 desvios padrão da média

Cerca de **99,7 %** de todos os valores ficam a 3 desvios padrão da média



Regras empíricas

Teorema de Chebyshev (TC) (pág. 82):

A proporção (ou fração) de qualquer conjunto de dados que se situa a K desvios padrão da média é *no mínimo*

$1 - 1/K^2$, onde K é um número positivo maior do que 1

Exemplos do Teorema de Chebyshev:

- Para $K = 2$, $1 - 1/2^2 = 3/4 = 0,75$. TC: Pelo menos 75% de todos os valores se localizam a 2 desvios padrão da média
- Para $K = 3$, $1 - 1/3^2 = 8/9 = 0,89$. TC: Pelo menos 89% de todos os valores se localizam a 3 desvios padrão da média

Capítulo 3 – Triola, 10^a. Ed.

Conteúdo

3.1 Visão Geral

3.2 Medidas de Centro (resumo)

3.3 Medidas de Variação

3.4 Medidas de Posição relativa

3.5 Análise Exploratória de Dados (AED)

3.4 Medidas de posição relativa

Escore-z: É o número de desvios padrão a que se situa determinado valor de x acima ou abaixo da média

amostra

$$z = \frac{x - \bar{x}}{s}$$

população

$$z = \frac{x - \mu}{\sigma}$$

Escores-z e valores não usuais:

Valores comuns: $-2 \leq z \leq 2$

Valores não-usuais: $z < -2$ ou $z > 2$

3.4 Medidas de posição relativa

Quartis e Percentis: Lembre-se da mediana, valor do meio, ou seja, 50% dos valores estão iguais ou menores do que a mediana.

A **mediana** divide os dados em duas partes iguais.

Os 3 **quartis** dividem os dados em **quatro** partes iguais:

São representados por Q_1 , Q_2 e Q_3 .

Os 99 **percentis** dividem os dados em **100** partes iguais.

São representados por $P_1, P_2, \dots, P_{98}, P_{99}$

3.4 Medidas de posição relativa

Note que: Q_1 separa os 25% inferiores dos 75% superiores

Q_2 separa os 50% inferiores dos 50% superiores

Q_3 separa os 75% inferiores dos 25% superiores

P_{13} separa os 13% inferiores dos 87% superiores

P_{75} separa os 75% inferiores dos 25% superiores

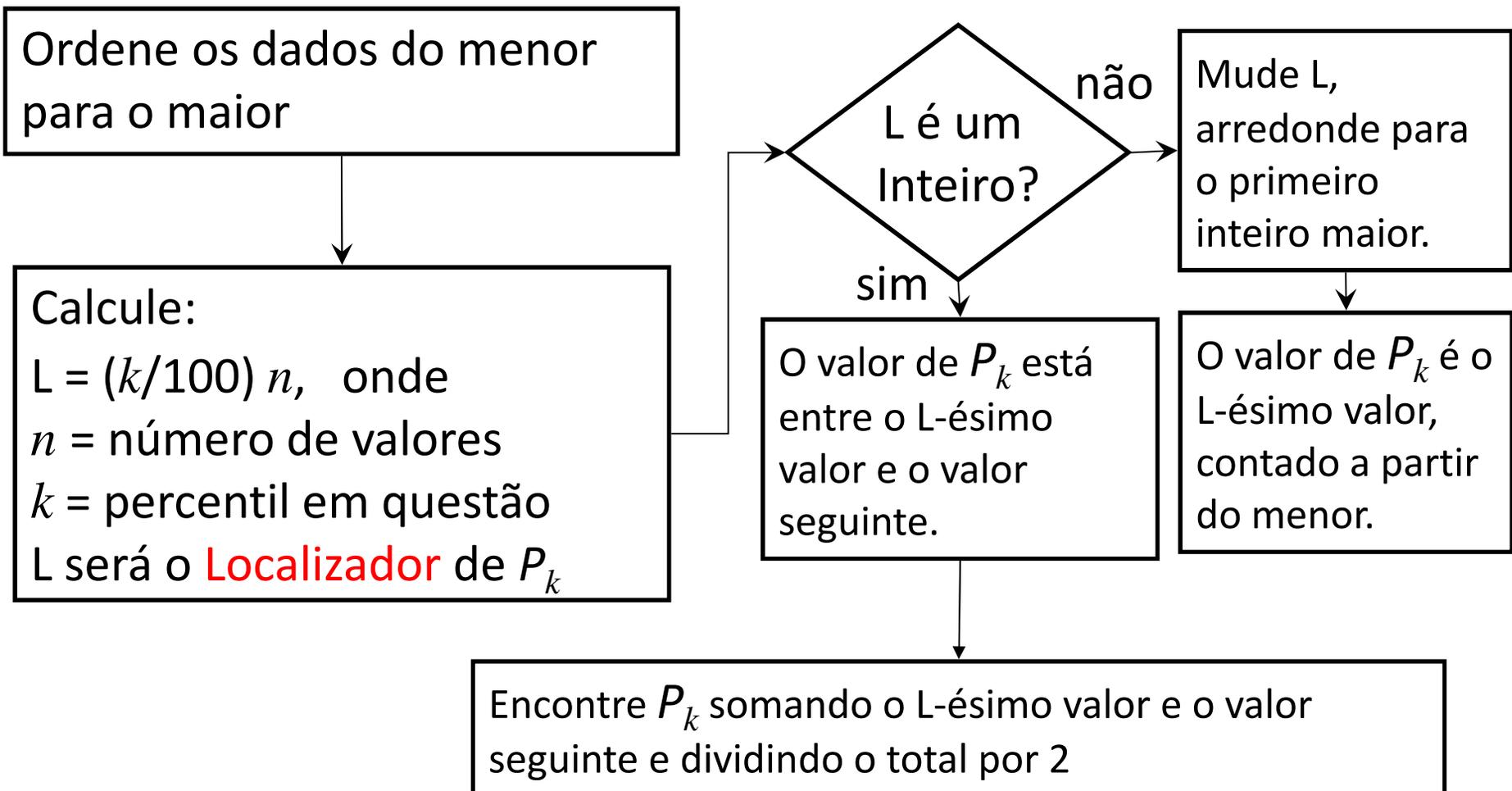
Assim,

$$Q_2 = P_{50} = \tilde{x} \quad (\text{mediana})$$

$$Q_1 = P_{25} \qquad Q_3 = P_{75}$$

3.4 Medidas de posição relativa

Procedimento para localizar P_k (o percentil k)



3.4 Medidas de posição relativa

Procedimento determinar qual é o percentil de um determinado valor x

$$\text{percentil do valor } x = \frac{\text{número de valores menores que } x}{\text{número total de valores}} \times 100$$

Capítulo 3 – Triola, 10^a. Ed.

Conteúdo

3.1 Visão Geral

3.2 Medidas de Centro (resumo)

3.3 Medidas de Variação

3.4 Medidas de Posição relativa

3.5 Análise Exploratória de Dados (AED)

3.5 Análise exploratória de dados

Uso de ferramentas estatísticas (gráficos, medidas de centro, medidas de variação) para investigar conjuntos de dados para compreender características importantes.

Diagrama de caixa (*Boxplot*) Um resumo dos cinco números

