Medidas de Posição ou Tendência Central

Medidas de Posição ou Tendência Central

- Fornece medidas que podem caracterizar o comportamento dos elementos de uma série;
- Possibilitando determinar se um valor está entre o maior e menor valor da série, ou se esta localizado no centro do conjunto de dados por exemplo.

Como definimos

- Média
- Mediana
- Moda
- Ponto Médio

Média Aritmética

- **a média aritmética de um conjunto de dados é o valor obtido somando-se todos os elementos do conjunto e dividindo-se a soma pelo número total de elementos**

$$\bar{x} = \frac{\sum x_i}{n}$$

onde $\bar{x}$ é a média aritmética, x_i os dados do conjunto amostral e n o número de valores.

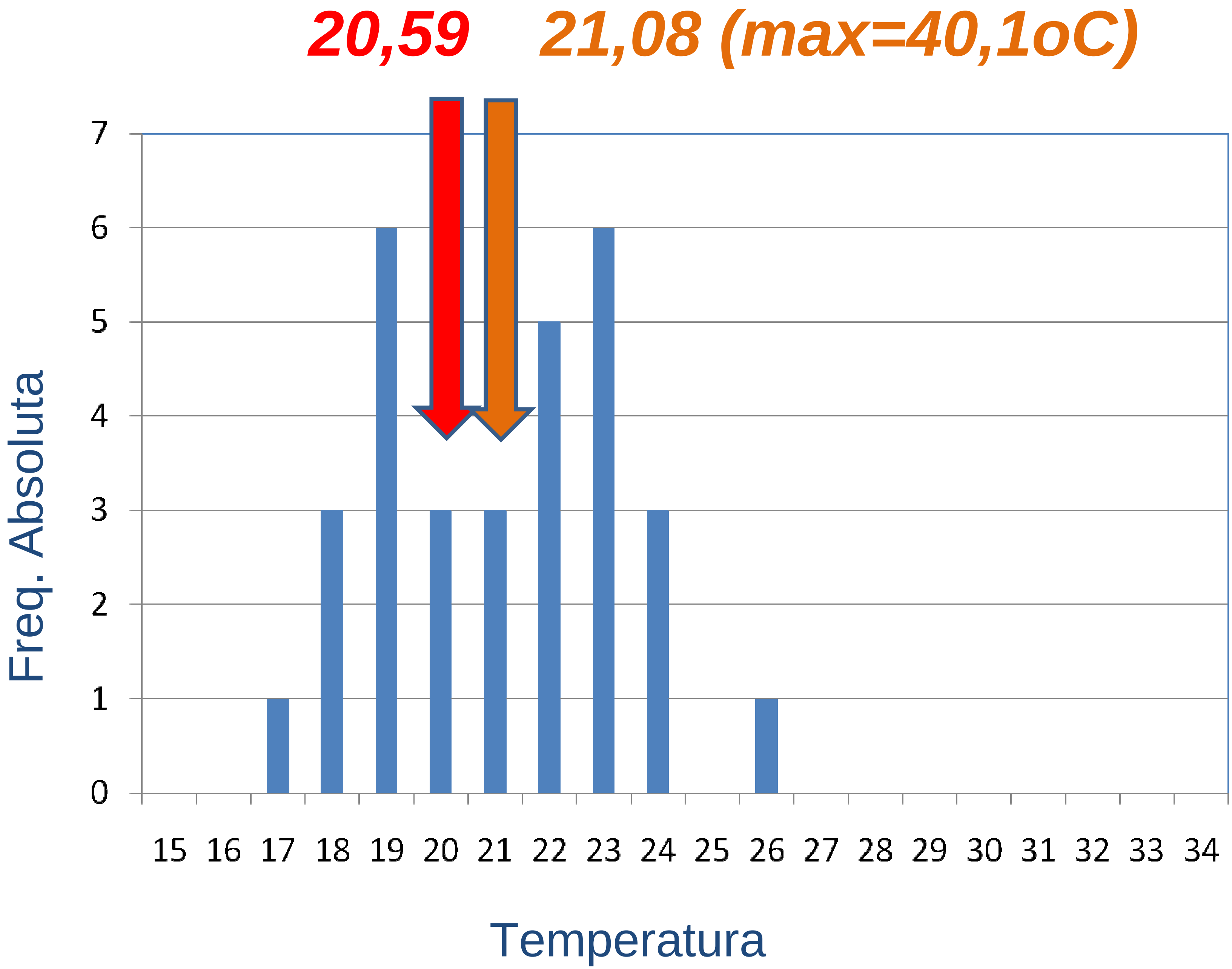
Exemplo: Temperatura média diária do mês de dezembro de 2004 da estação do IAG.

Dia	Temperatura (°C)	Dia	Temperatura (°C)
1	18,9	17	21,5
2	18,7	18	20,8
3	18,4	19	22,4
4	23,2	20	23,7
5	22,3	21	18,3
6	22	22	16,1
7	22,4	23	17,2
8	23	24	19,8
9	20,9	25	22,6
10	18,3	26	21,2
11	17,5	27	21,2
12	18	28	20,1
13	19,1	29	21,4
14	18,9	30	22,2
15	20	31	23,2
16	25,1		

A média aritmética calculada para a Temperatura média diária do mês de Dezembro de 2004

$$\bar{x} = \frac{18,9 + 18,7 + ... + 22,2 + 23,2}{31}$$

$$\bar{x} = 20,59^{0}\ C$$



20,59
21,08 (max=40,1oC)
Freq. Absoluta
7
6
5
4
3
2
1
0
15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34
Temperatura

Média Harmônica

- costuma ser usada como medida de tendência central para conjuntos de dados que consistem em taxas de variação, como por exemplo velocidades.

$$\bar{x} = \frac{n}{\sum \frac{1}{x_i}}$$

A média harmônica calculada para a Temperatura média diária do mês de Dezembro de 2004

$$\bar{x} = \frac{31}{\frac{1}{18,9} + \frac{1}{18,7} + \ldots + \frac{1}{22,2} + \frac{1}{23,2}}$$

$$\bar{x} = 20,36^{0}\ C$$

Média Geométrica

- é usada na administração e na economia para achar taxas médias de variação, de crescimento, ou razões médias.
- Dados n valores (todos positivos), a média geométrica é a raiz n^ma^ do seu produto (Triola, 1998)

$$\bar{x} = \sqrt[n]{x_1 * x_2 * x_3 \ldots\ldots\ldots\ldots\ldots * x_{n-2} * x_{n-1} * x_n}$$

A média geométrica calculada para a Temperatura média diária do mês de Dezembro de 2004

$$\bar{x} = \sqrt[31]{18,9 * 18,7 * \ldots * 22,2 * 23,2}$$

$$\bar{x} = 20,48^{0}\ C$$

Média Quadrática

- é utilizada em geral em experimentos físicos;
- Em sistemas de distribuição de energia, por exemplo, as tensões e correntes são em geral dadas em termos de sua média quadrática;
- Obtém-se a média quadrática de um conjunto de valores elevando-se cada um ao quadrado, somando-se os resultados, dividindo-se o total pelo número n de valores e tomando-se a raiz quadrada do resultado (Triola, 1998).

$$\bar{x} = \sqrt{\frac{\sum x_i^2}{n}}$$

A média quadrática calculada para a Temperatura média diária do mês de Dezembro de 2004

$$\bar{x} = \sqrt{\frac{(18,9)^2 + (18,7)^2 + \ldots + (22,2)^2 + (23,2)^2}{31}}$$

$$\bar{x} = 20,71^0\ C$$

Mediana

- A mediana é o elemento que ocupa a posição central de uma série de dados. Para encontrá-la os dados devem estar dispostos em ordem crescente ou decrescente;
- Se a série tiver um número ímpar de dados o valor que estiver ocupando o meio da série será a mediana;
- Se tiver um número par de dados deve-se extrair a média aritmética dos dois valores centrais, uma vez que, o valor correspondente a mediana acha-se entre eles.

Ordenando a Tabela de Temperatura Média do Ar temos:

1	16,1
2	17,2
3	17,5
4	18
5	18,3
6	18,3
7	18,4
8	18,7
9	18,9
10	18,9
11	19,1
12	19,8
13	20
14	20,1
15	20,8

16	**20,9**

17	21,2
18	21,2
19	21,4
20	21,5
21	22
22	22,2
23	22,3
24	22,4
25	22,4
26	22,6
27	23
28	23,2
29	23,2
30	23,7
31	25,1

Moda

- A moda é o valor que ocorre com maior freqüência em uma série de dados.
- Pode ser identificada apenas observando-se a série nos casos de dados não agrupados.
- Quando a série possuir dois valores com a mesma freqüência máxima, cada um deles é uma moda, e o conjunto diz-se bimodal.

Moda

- Se mais de dois valores ocorrerem com a mesma freqüência máxima, o conjunto é multimodal.
- Quando nenhum valor é repetido, o conjunto não tem moda.

Série de Temp. do Ar ordenada e com o número de ocorrências

Temperatura	Freqüências
16,1	1
17,2	1
17,5	1
18	1
18,3	2
18,4	1
18,7	1
18,9	2
19,1	1
19,8	1
20	1
20,1	1
20,8	1

Temperatura	Freqüências
20,9	1
21,2	2
21,4	1
21,5	1
22	1
22,2	1
22,3	1
22,4	2
22,6	1
23	1
23,2	2
23,7	1
25,1	1

Moda: Multimodal

Temperatura	Freqüências
16,1	1
17,2	1
17,5	1
18	1
18,3	**2**
18,4	1
18,7	1
18,9	**2**
19,1	1
19,8	1
20	1
20,1	1
20,8	1

Temperatura	Freqüências
20,9	1
21,2	**2**
21,4	1
21,5	1
22	1
22,2	1
22,3	1
22,4	**2**
22,6	1
23	1
23,2	**2**
23,7	1
25,1	1

Ponto Médio

- **O ponto médio é o valor que está a meio caminho entre o maior e o menor valor da série de dados:**

$$PM = \frac{\text{maior valor} + \text{menor valor}}{2}$$

O ponto médio para a Temperatura média diária do mês de Dezembro de 2004

$$PM = \frac{16{,}1 + 25{,}1}{2}$$

PM = 20,6º C

Medidas de Dispersão ou de Variabilidade

Média/Variabilidade

- **X: 70, 70, 70, 70, 70** $\bar{x} = \frac{\sum x_i}{n} = \frac{350}{5} = 70$

- **Y: 68, 69, 70, 71, 72** $\bar{y} = \frac{\sum y_i}{n} = \frac{350}{5} = 70$

- **Z: 5, 15, 50, 120, 160** $\bar{z} = \frac{\sum z_i}{n} = \frac{350}{5} = 70$

Analisando a Dispersão ou Variabilidade

- Amplitude Total
- Desvio Padrão
- Variância
- Assimetria

Amplitude Total

- A amplitude total de um conjunto de dados é a diferença entre o maior e o menor valor deste.

$$AT = x_{máx} - x_{mín}$$

- Quanto maior a amplitude total de um conjunto de dados, maior é a dispersão ou variabilidade dos valores.

Para a série de Temperatura do Ar

AT = 25,1 – 16,1 = 9º C

Desvio-Padrão

- uma medida da magnitude do espalhamento ou dispersão dos dados em relação à média da série.

desvio-padrão amostral (s) é

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

onde x_i é cada elemento do conjunto de dados, $\bar{x}$ é a média do conjunto e n é o número total de elementos deste.

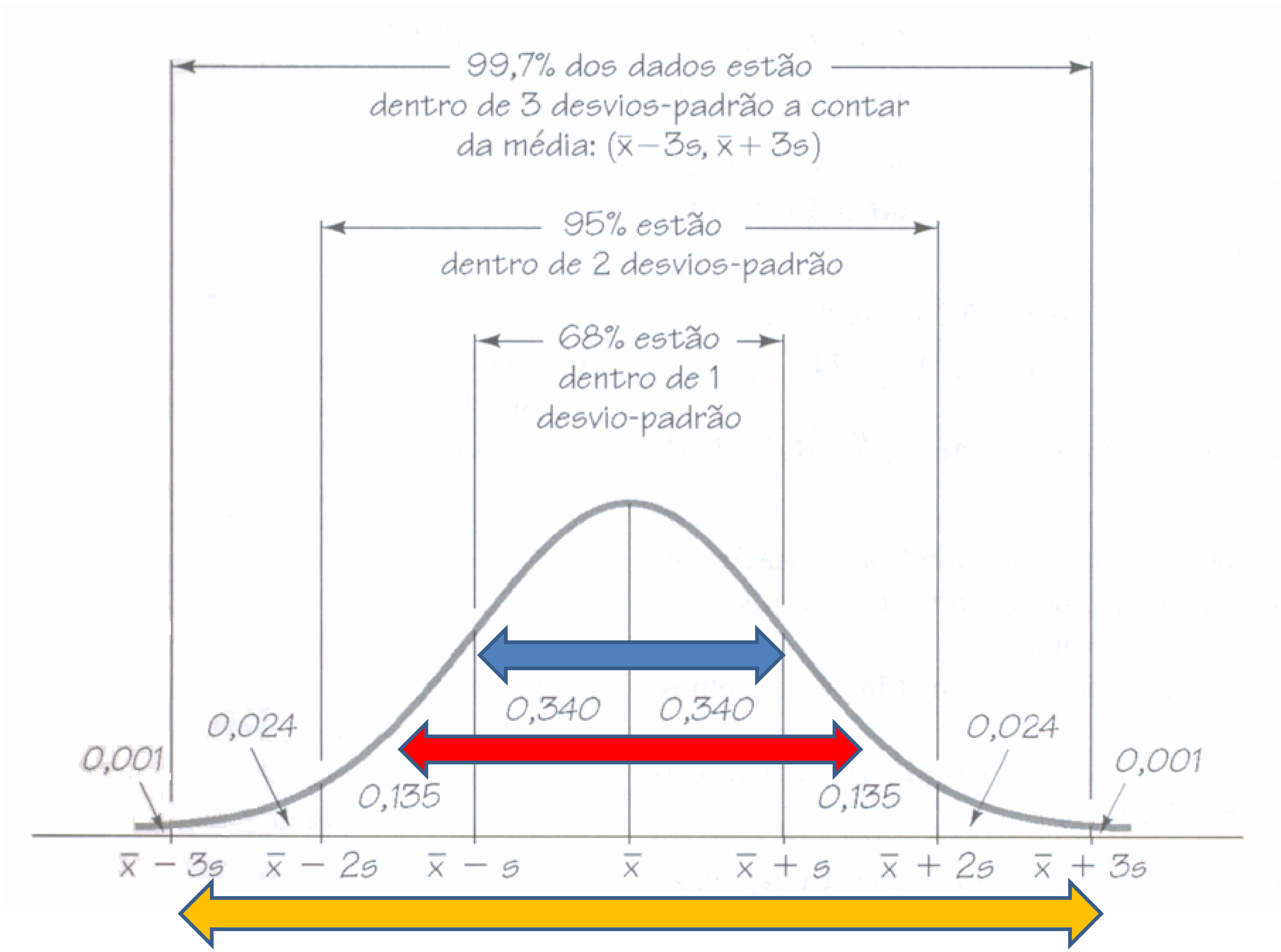
o desvio-padrão populacional (σ)

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \mu\right)^2}{N}}$$

- onde x_i é cada elemento da população, μ e N são respectivamente a média e o número total de elementos da população.

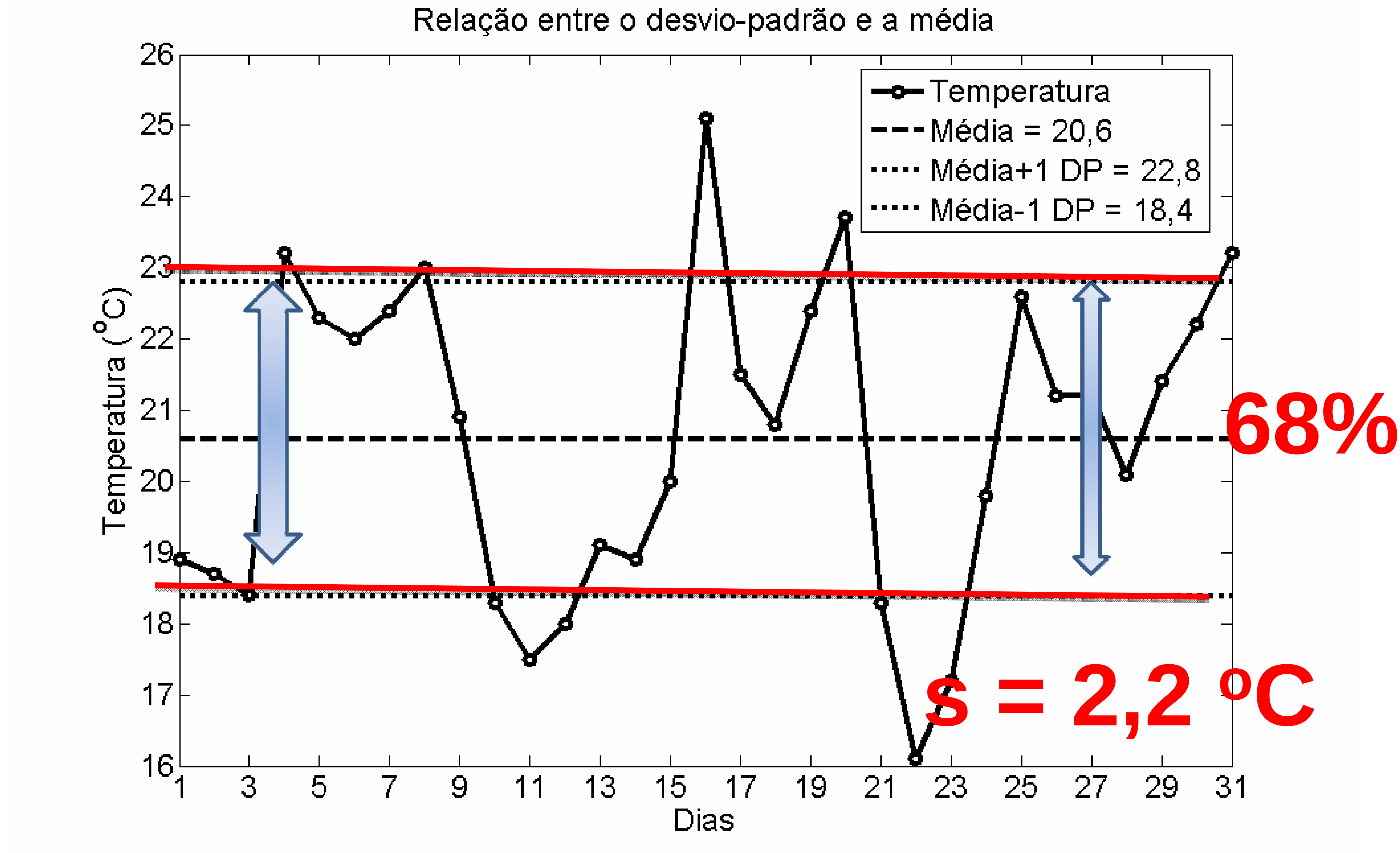
Desvio Padrão

- Uma regra que auxilia na interpretação do valor de um desvio-padrão é a regra empírica, aplicável somente a conjuntos de dados aproximadamente em forma de sino.
- ***a. cerca de 68% dos valores estão a menos de 1 desvio-padrão a contar da média;***
- ***b. cerca de 95% dos valores estão a menos de 2 desvios-padrão a contar da média;***
- ***c. cerca de 99,7% dos valores estão a menos de 3 desvios-padrão a contar da média.***



99,7% dos dados estão dentro de 3 desvios-padrão a contar da média: ($\bar{x}-3s$, $\bar{x}+3s$)
95% estão dentro de 2 desvios-padrão
68% estão dentro de 1 desvio-padrão
0,340
0,340
0,024
0,024
0,001
0,001
0,135
0,135
$\bar{x}-3s$
$\bar{x}-2s$
$\bar{x}-s$
$\bar{x}$
$\bar{x}+s$
$\bar{x}+2s$
$\bar{x}+3s$

Temperatura do Ar Média Diária do Mês de Dezembro de 2004.



Variância

- A variância é uma medida estatística da dispersão dos dados em torno da média de um conjunto de dados, é o quadrado do desvio-padrão.

Amostral

$$s^2 = \frac{\sum (x_i - \overline{x})^2}{n-1}$$

Populacional

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

Qual outra função do Desvio Padrão e Variância

- Identificação de “outliers”
  - Valores que estão muito distantes da média, ou seja, podem ser valores estranhos ou erros.
  - podem ser erros de coleta/digitação ou um real desvio da amostragem
  - Portanto devemos sempre analisar estes outliers antes de descartá-los.
- Verificação de desvios na amostra

Identificando um Outlier

- De acordo com a regra de Tchebichev
- Valores fora do intervalo de ± 2 s (desvio padrão) podem ser outliers e devem ser analisados

→ fora deste intervalo, **possível dado estranho**

$$\left[\bar{x} - 2s; \bar{x} + 2s\right]$$

Desvio da Amostra

- Escore Padronizado

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

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

Amostral Populacional

- Número de desvios-padrão pelo qual um valor dista da média (para mais ou para menos)

Exemplo

- A temperatura média do ar para o mês de Dezembro de 2004, $\overline{x}$ = 20,59 ^{o}C enquanto que o desvio padrão, s = 2,2 ^{o}C.
- ***Suponha que em um dado dia de Novembro, a Temperatura do ar média diária foi de 35 ^{o}C. Seria esta temperatura excepcionalmente quente para este mês?***

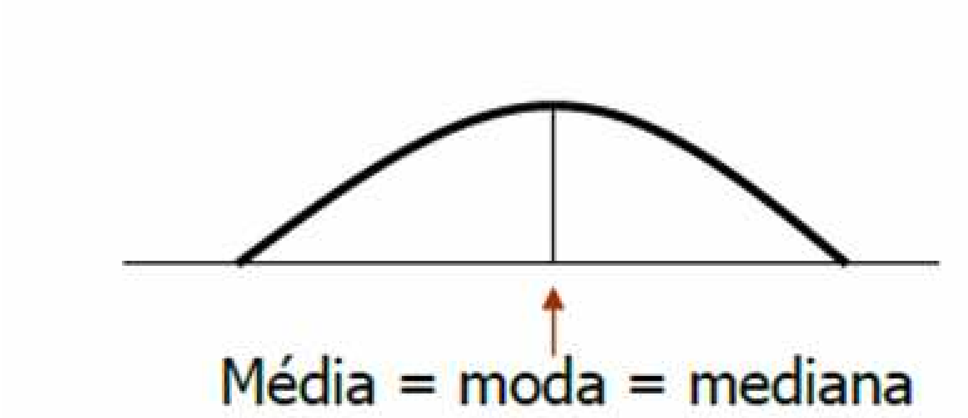
Calculando o escore Z

$$z = \frac{(35{,}0 - 20{,}59)}{2{,}2} = 7{,}20$$

- *Este resultado indica que a temperatura do ar média daquele dia está a 7,20 desvios-padrão acima da média da amostra.*
- *Assumindo que este era uma medida real, temos que este dia foi realmente quente para aquele mês.*

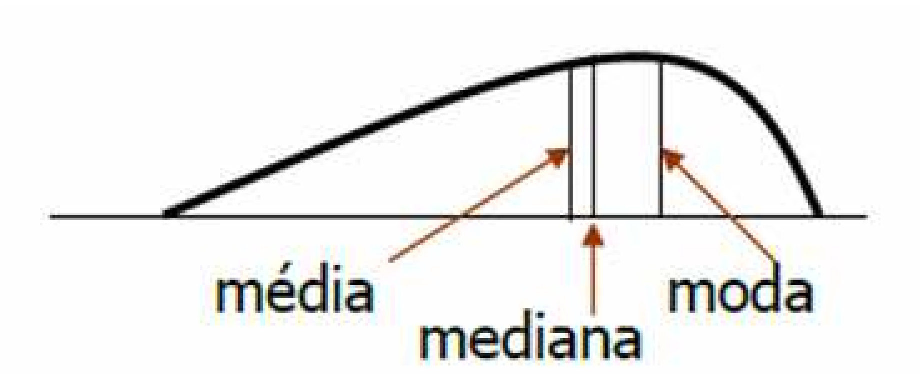
Assimetria

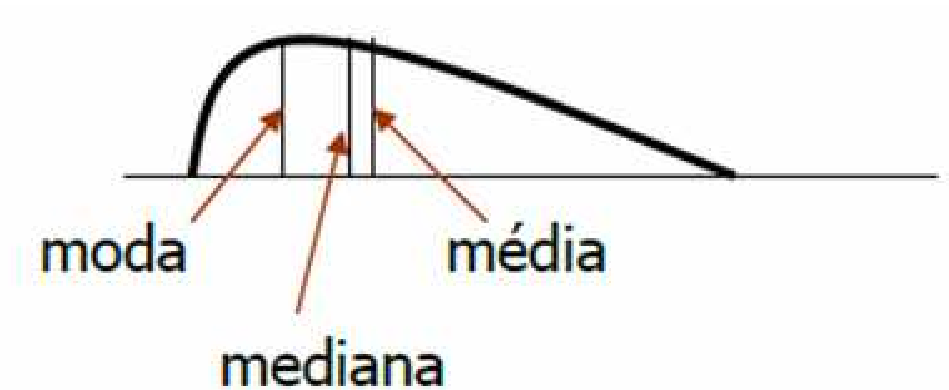
- grau de deformação de uma curva de freqüências.
  - Simétrica
  - Assimétrica positiva
  - Assimétrica negativa



Simétrica

$$\bar{x} = Me = Mo$$





Assimétrica Negativa

$$\bar{x} < Me < Mo$$

Assimétrica Positiva

$$\bar{x} > Me > Mo$$

Parâmetros da Assimetria

- coeficiente de assimetria (A)

$$A = \frac{\overline{x} - Mo}{s}$$

Onde Mo é a moda

- Se for difícil determinar a moda, o coeficiente de assimetria é obtido com boa aproximação:

$$A = \frac{3(\overline{x} - Me)}{s}$$

Onde Me é a mediana

- Porém a medida de assimetria mais utilizada é dada pelo terceiro momento (m_3), centrado na média.

$$A = \frac{m_3}{s^3}$$

Onde m_3 é dado por:

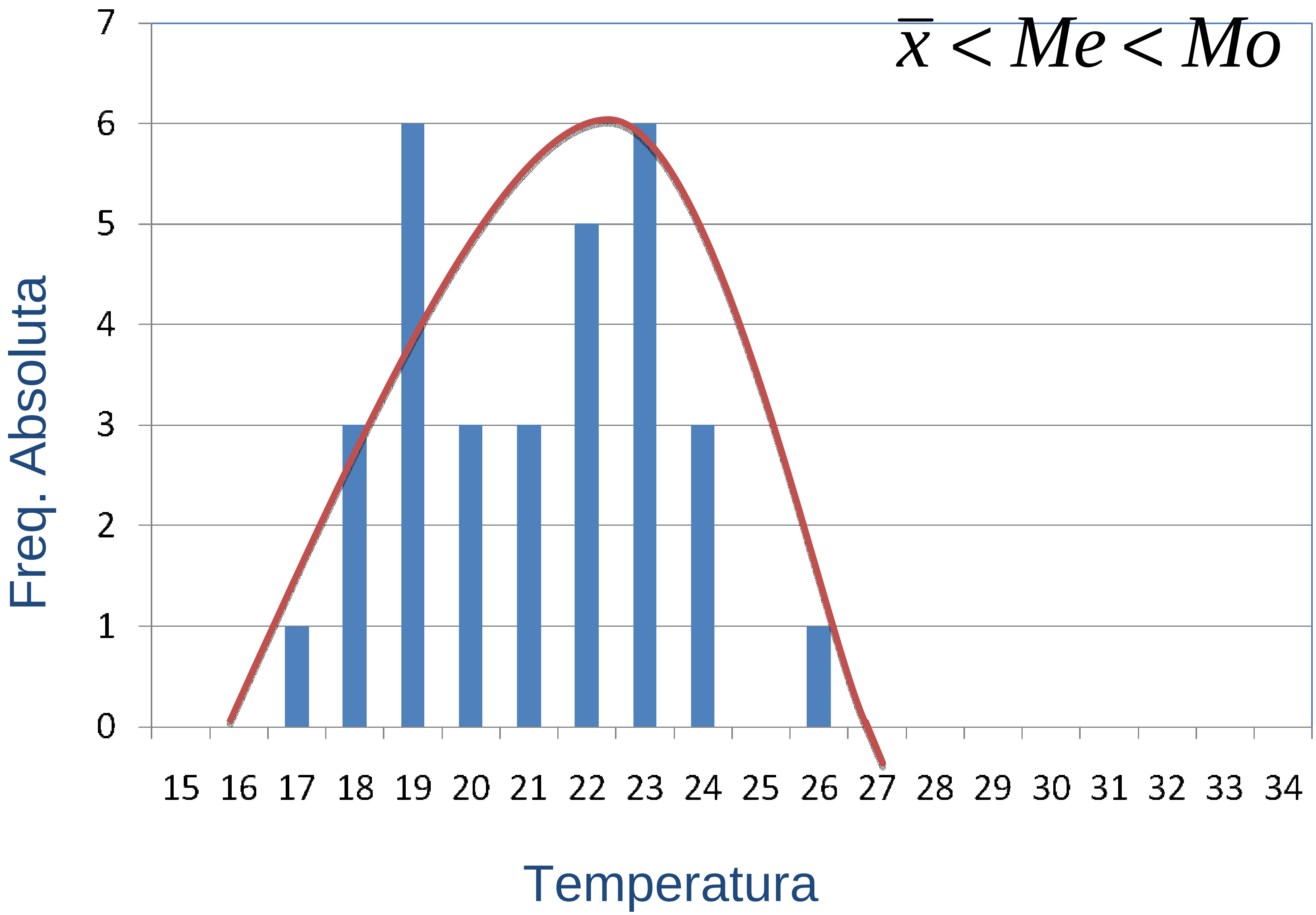
$$m_3 = \frac{\Sigma \left(x_i - \overline{x}\right)^3}{n}$$

Classificação

- A = 0 distribuição simétrica
- A > 0 distribuição assimétrica positiva
- A < 0 distribuição assimetria negativa.

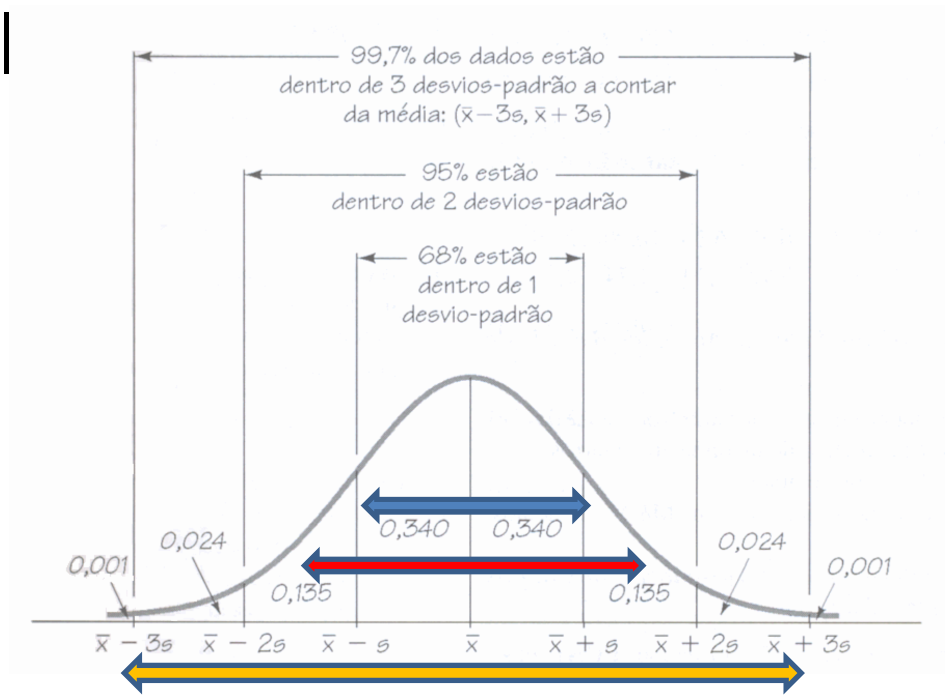
A = - 0,08 → Assimetria Negativa

$$\bar{x} < Me < Mo$$

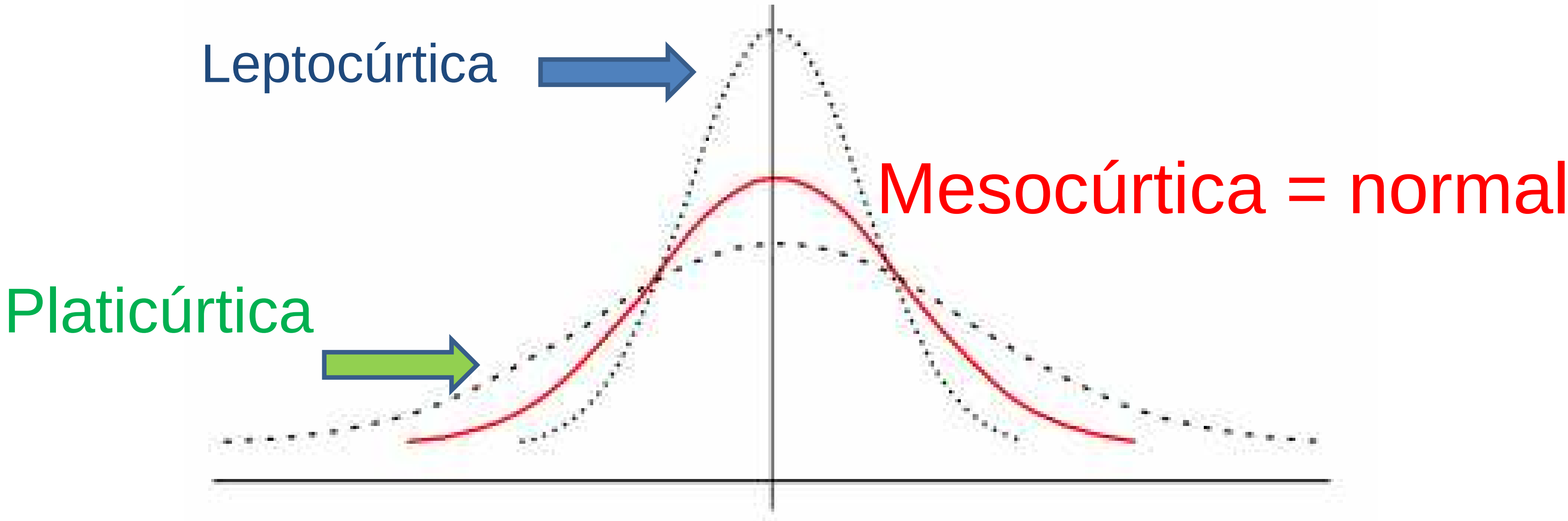


Curtose

- A curtose é o grau de achatamento de uma distribuição em relação a uma distribuição padrão, denominada curva normal



Tipos



- A curtose (C) é definida pelo quarto momento (m_4) dividido pelo o desvio-padrão da série elevado a quarta potência (s^4):

$$C = \frac{m_4}{s^4}$$

onde m_4 é $$m_4 = \frac{\Sigma\,(x_i - \bar{x})^4}{n}$$

Valores de Curtose

- C = 3 a curtose é denominada mesocúrtica (*curva normal*);
- C > 3, a curtose é denominada leptocúrtica (*a curva mais fechada que a curva normal*)
- C < 3, a curtose é denominada platicúrtica (*a curva é mais achatada que a curva normal*)

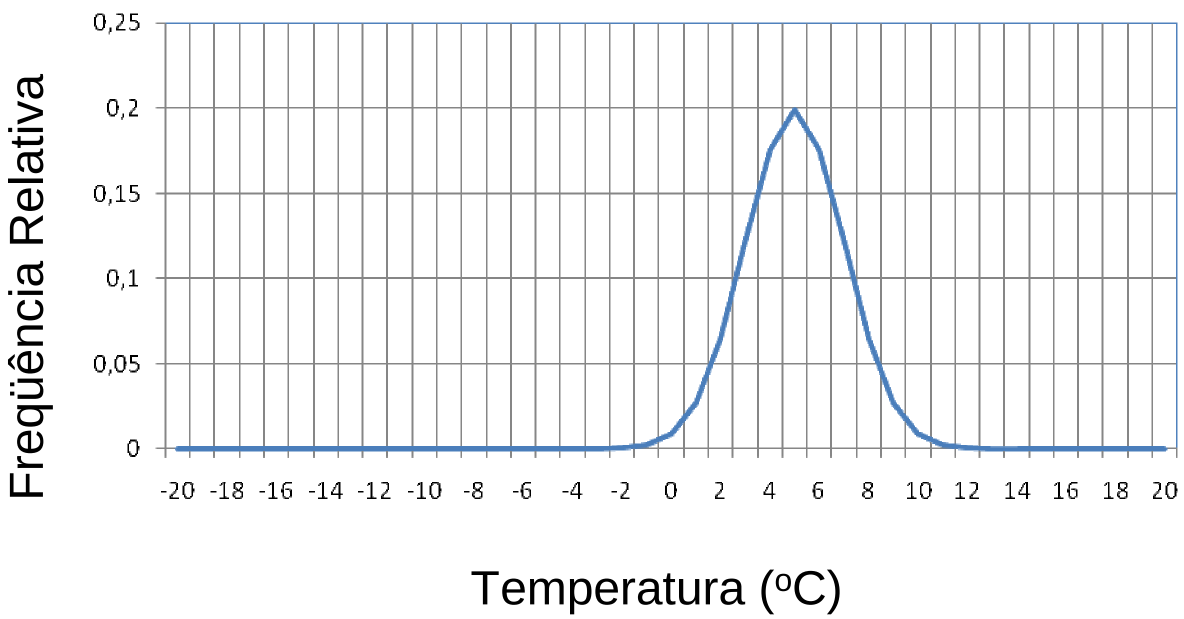
A curtose calculada para a Temperatura do ar média diária para o mês de Dezembro é $C = 2,2$, portanto $C<3$ e a curva de freqüência é mais achatada que a curva normal.

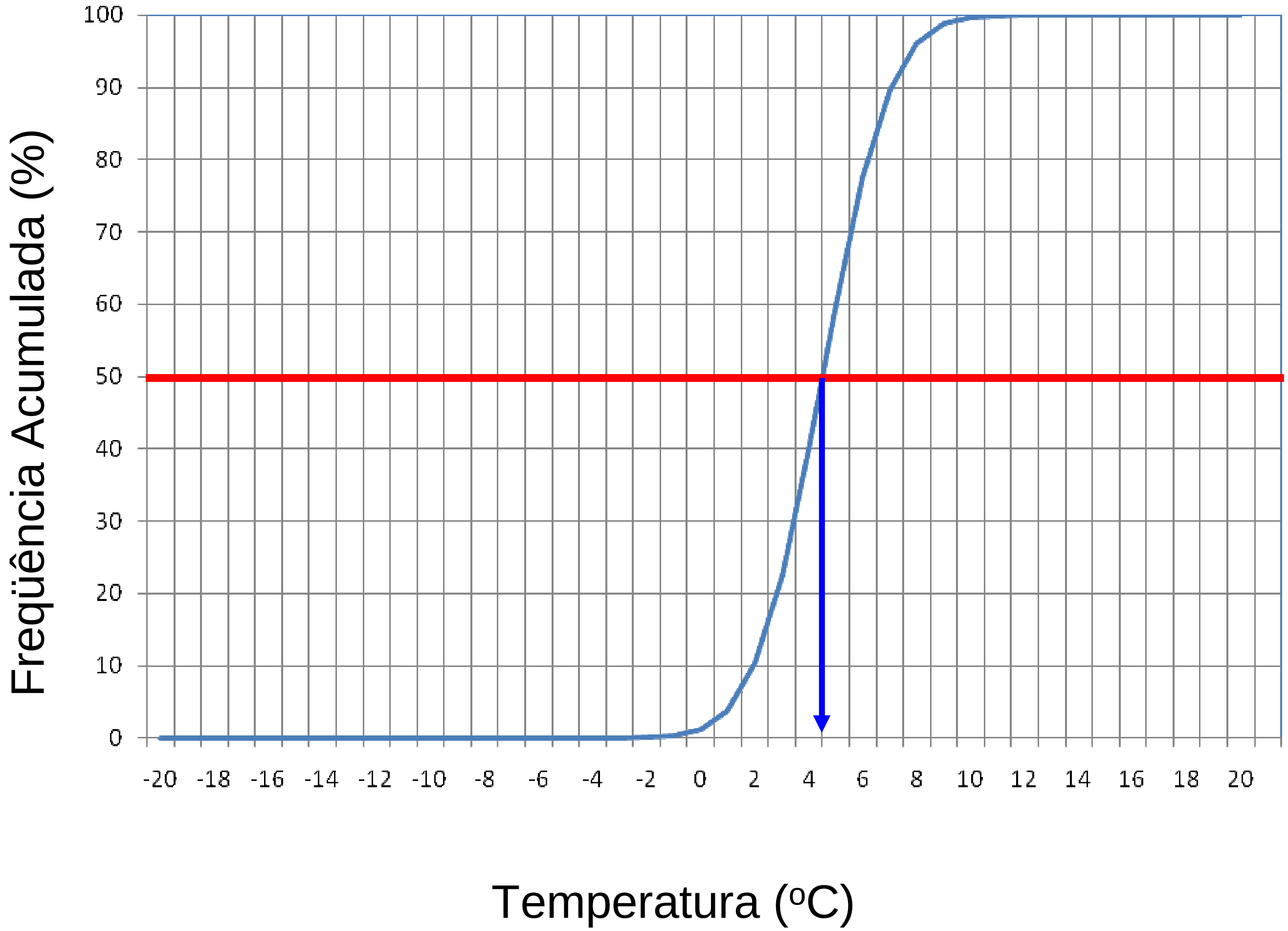
Separatrizes

- Mediana
- Quantis
- fractis

Mediana

- Separa a série em dois grupos que apresentam o mesmo número de valores.
- Ou ainda, representa a posição com 50% da distribuição.




Freqüência Acumulada (%)
100
90
80
70
60
50
40
30
20
10
0
-20
-18
-16
-14
-12
-10
-8
-6
-4
-2
0
2
4
6
8
10
12
14
16
18
20
Temperatura (oC)

Quantil

- Nome genérico para outras medidas, como as que dividem o conjunto de dados em 4, 10 ou 100 partes, por exemplo.
- Estas separações são denominadas de quartil, decil e percentil, respectivamente.

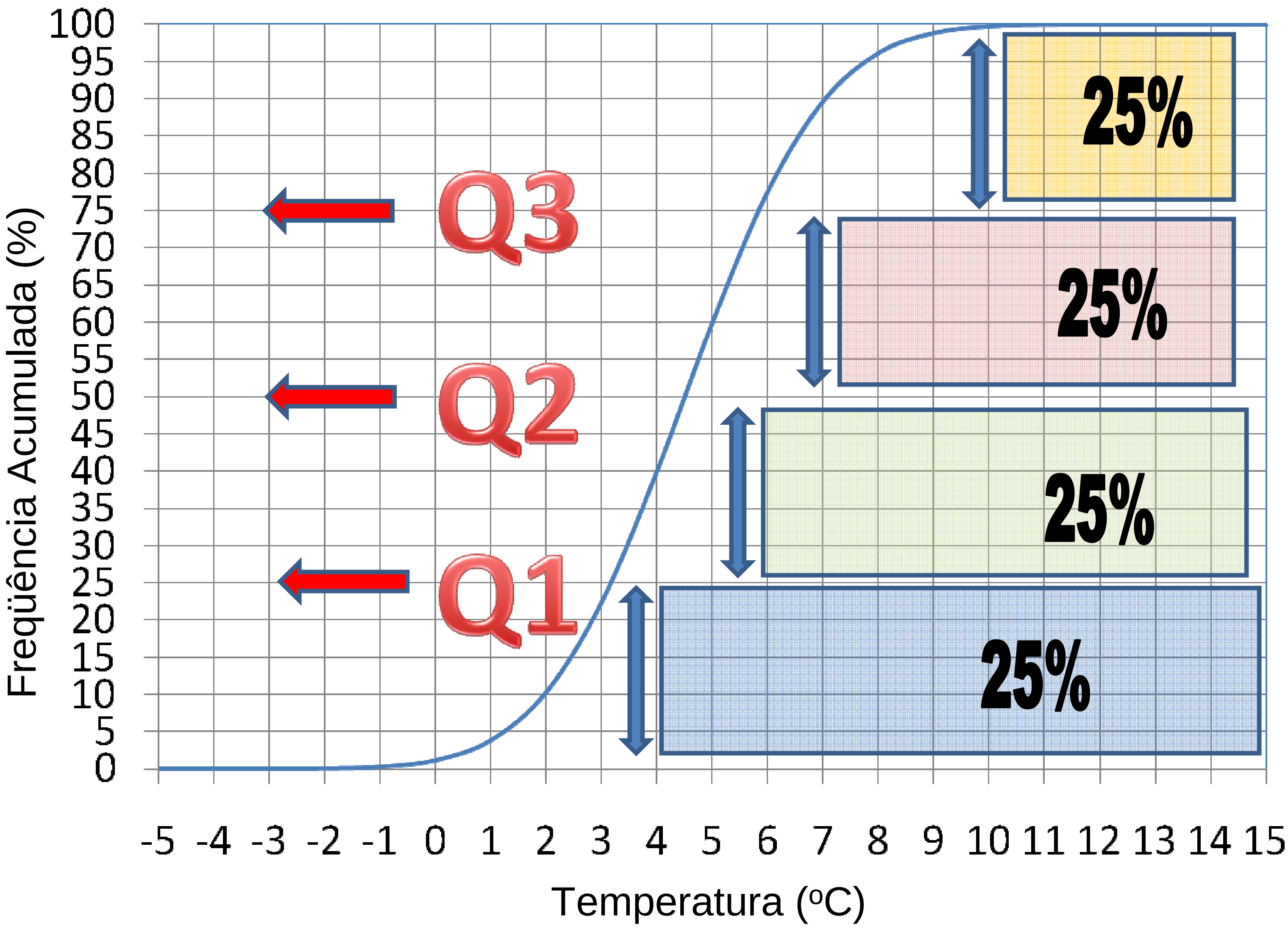
Quantil

- Um quantil amostral $\mathbf{q_p}$ é um número tendo a mesma unidade que o dado, o qual excede a proporção do dado dada pelo subscrito **p**, com $0 \leq p \leq 1$.
- Em outras palavras, o quantil amostral $\mathbf{q_p}$ pode ser interpretado aproximadamente como aquele valor do dado que ***excede*** um membro escolhido aleatoriamente do conjunto de dado, com *probabilidade p*.

Quartil

- Os três quartis Q_1, Q_2 e Q_3 dividem o conjunto dos dados em quatro subconjuntos de tal forma que:

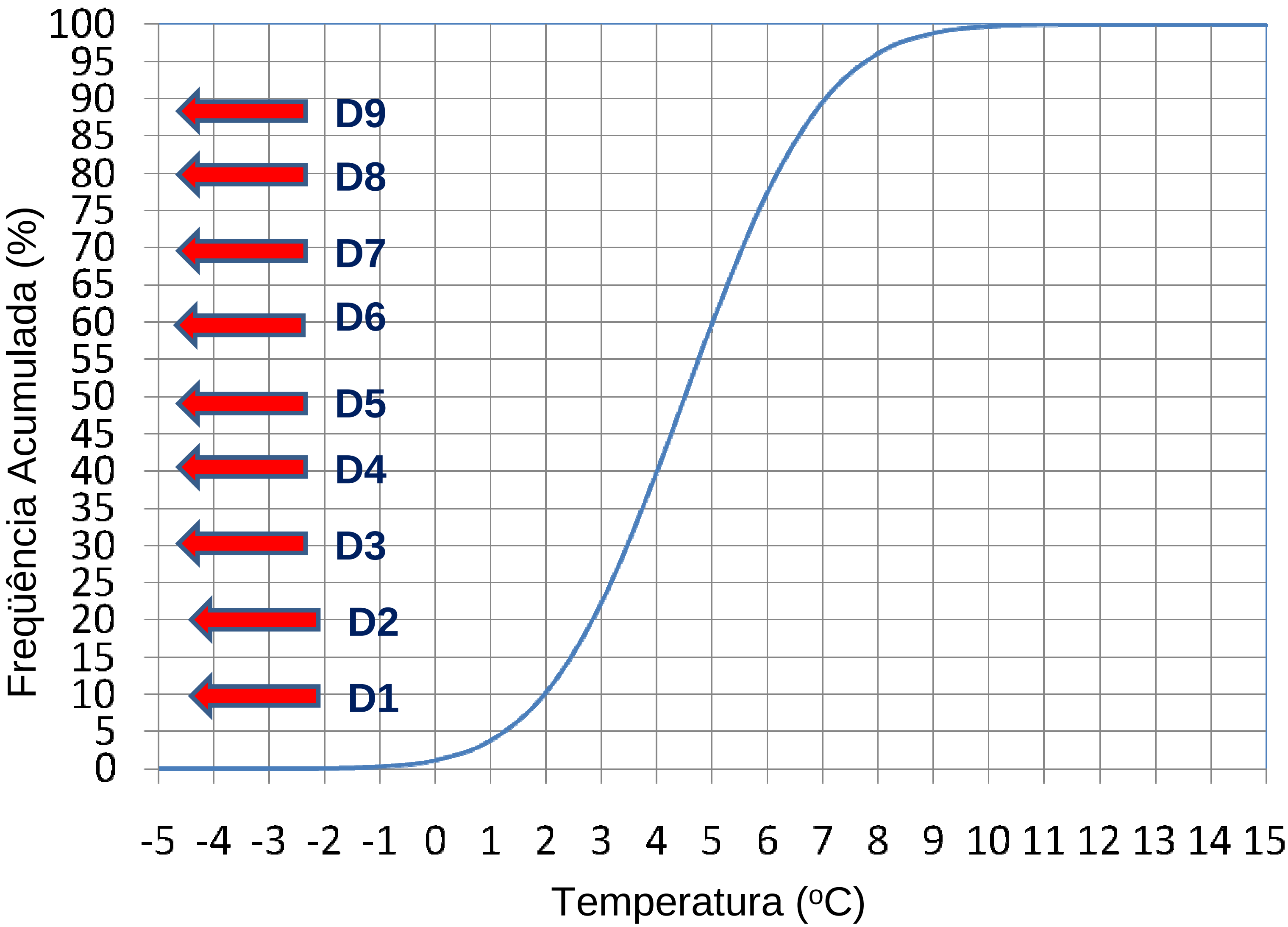
- *25% dos elementos situam-se abaixo do Q_1;*
- *25% entre Q_1 e Q_2;*
- *25% entre Q_2 e Q_3 e;*
- *25% acima de Q_3, sendo que Q_2 corresponde a mediana.*



Freqüência Acumulada (%)
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
Q3
Q2
Q1
25%
25%
25%
25%
-5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Temperatura (oC)

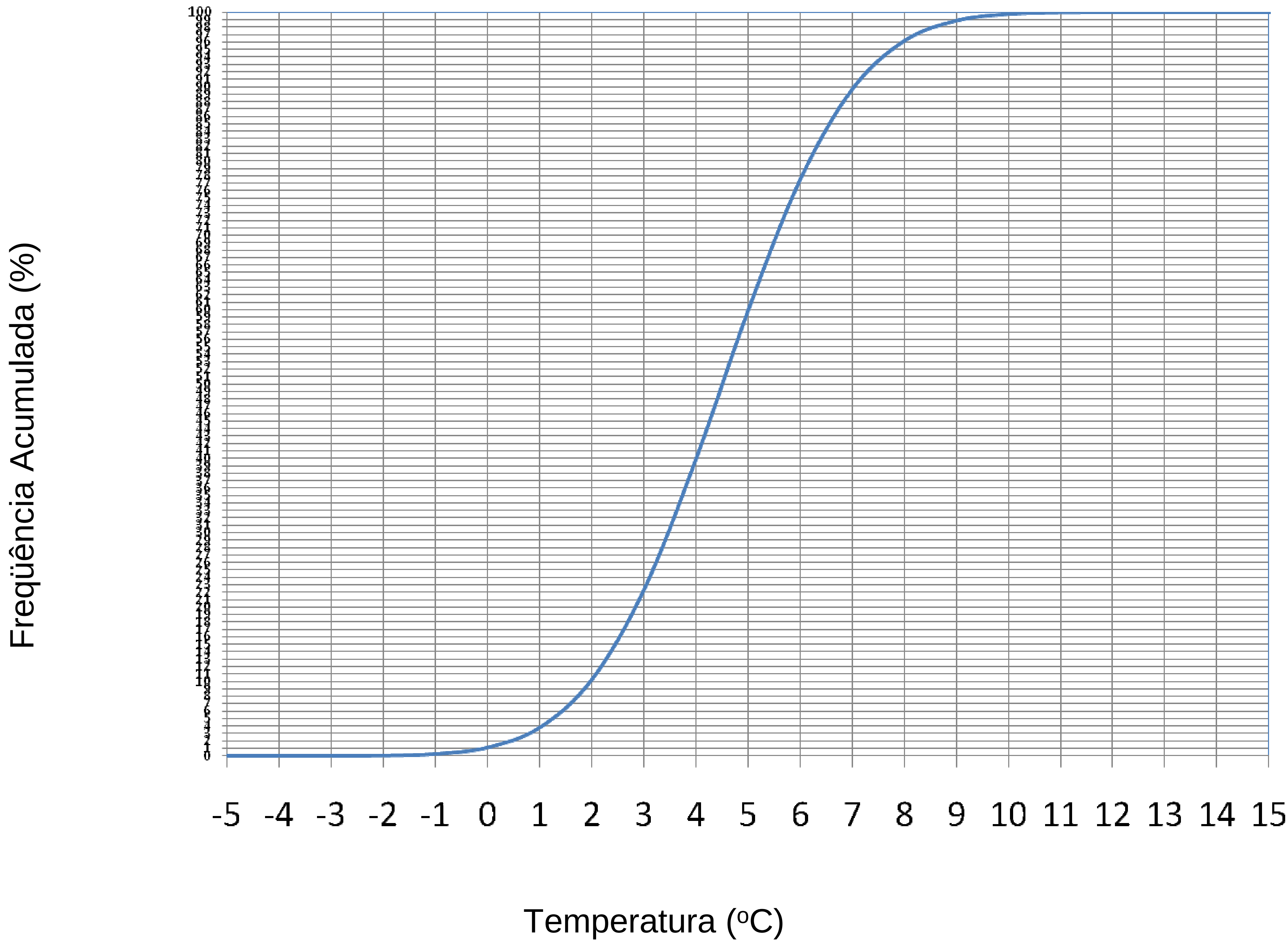
Decis

- Os decis dividem o conjunto de dados em 10 partes iguais.
- Os nove decis D_1, D_2, D_3,..., D_9 são tais que 10% dos elementos situam-se abaixo de D_1, 10% entre D_1 e D_2 e assim por diante.
- A mediana é o quinto decil.


Freqüência Acumulada (%)
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
D9
D8
D7
D6
D5
D4
D3
D2
D1
-5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Temperatura (oC)

Percentil

- Os percentis dividem o conjunto dos dados ordenados em 100 partes iguais.
- A mediana é o qüinquagésimo percentil.



Freqüência Acumulada (%)
Temperatura (ºC)
-5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
0 10 100

Procedimento para obtenção dos quantis **(Xavier et al., 2002):**

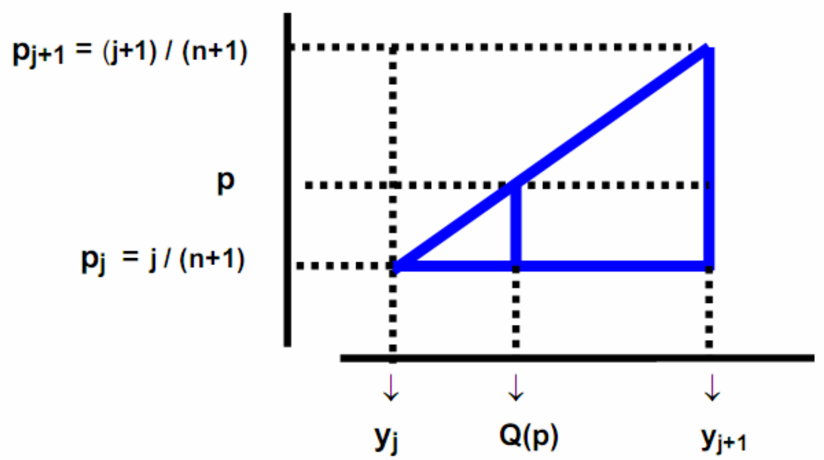
- 1) Dispor os dados em ordem crescente;
- 2) Colocar um número de ordem para cada valor ($i=1, ..., i=N$);
- 3) Para cada valor determinar a ordem quantílica: $P_i=i/(N+1)$, onde N é o número de elementos da série;
- 4) Finalmente, para calcular o quantil Q(P) para uma ordem quantílica P_i qualquer, segue-se:

  a) se P coincidir com algum P_i já obtido, então:

  $Q(P)=Q(P_i)=y_i$

Procedimento para obtenção dos quantis **(Xavier et al., 2002):**

b) se P não coincidir, haverá um índice i tal que $P_i < P < P_{i+1}$, onde Q(P) será obtido por interpolação, onde:



$$Q(P) = v_i + \left[\frac{P - P_i}{P_{i+1} - P_i}\right] x \left(v_{i+1} - v_i\right)$$

Exemplo:

Mês	*Temperatura Julho-Agosto 2002*
1	*23.764*
2	*23.614*
3	*21.827*
4	*22.196*
5	*22.023*
6	*24.438*
7	*24.675*
8	*24.244*
9	*24.749*
10	*25.611*
11	*25.002*
12	*24.179*

Mês	Temperatura Julho-Agosto 2002	N ordem	Temperatura Ordenada
1	23.764	1	25.611
2	23.614	2	25.002
3	21.827	3	24.749
4	22.196	4	24.675
5	22.023	5	24.438
6	24.438	6	24.244
7	24.675	7	24.179
8	24.244	8	23.764
9	24.749	9	23.614
10	25.611	10	22.196
11	25.002	11	22.023
12	24.179	12	21.827

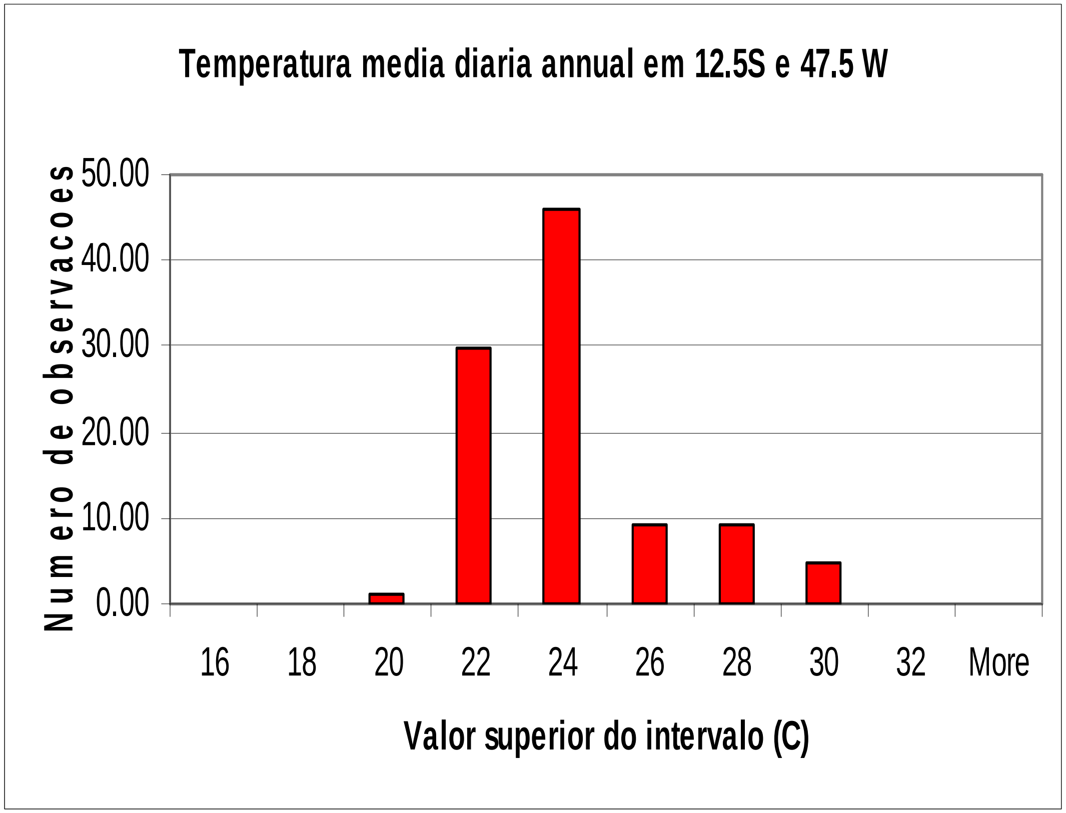
1.) Dispor os dados em ordem crescente;

2.) Colocar um número de ordem para cada valor (i=1, ..., i=N);

Mês	Temperatura Julho-Agosto 2002	Temperatura ordenada	Pi	Percentil(%)
1	23.764	25.611	1/(1+12)	7,7
2	23.614	25.002	2/(1+12)	15,4
3	21.827	24.749	3/(1+12)	23,1
4	22.196	24.675	4/(1+12)	30,8
5	22.023	24.438	5/(1+12)	38,5
6	24.438	24.244	6/(1+12)	46,2
7	24.675	24.179	7/(1+12)	53,8
8	24.244	23.764	8/(1+12)	61,5
9	24.749	23.614	9/(1+12)	69,2
10	25.611	22.196	10/(1+12)	76,9
11	25.002	22.023	11/(1+12)	84,6
12	24.179	21.827	12/(1+12)	92,3

3.) Para cada valor determinar a ordem quantílica: P_i=i/(N+1), onde N é o número de elementos da série;

Exemplo 2



Classe Temp	*Freq Absoluta*	Freq.Rel. (%)
16.01- 18.00	0	0.00
18.01- 20.00	17	0.97
20.01- 22.00	523	29.85
22.01- 24.00	806	46.00
24.01- 26.00	159	9.08
26.01- 28.00	160	9.13
28.01- 30.00	84	4.79
30.01- 32.00	3	0.17
total	1752	100%

Classe Temp	*Freq Absoluta*	Acumulada	Freq.Rel. (%)	Freq. Acumulada (%)
16.01- 18.00	0	0	0.00	0
18.01- 20.00	17	17	0.97	0,97
20.01- 22.00	523	540	29.85	30,82
22.01- 24.00	806	1346	46.00	76,83
24.01- 26.00	159	1505	9.08	85,90
26.01- 28.00	160	1665	9.13	95,03
28.01- 30.00	84	1749	4.79	99,83
30.01- 32.00	3	1752	0.17	100,00
total	1752		100%	

Quantils	Temperatura
0,25	**20,61**
0,50	
0,75	
0,90	

$$Q(P) = v_i + \left[\frac{P - P_i}{P_{i+1} - P_i} \right] x \left(v_{i+1} - v_i \right)$$

$$Q(25\%) = 20{,}01 + \left[\frac{25{,}0 - 0{,}97}{30{,}82 - 0{,}97} \right] x \left(22{,}0 - 20{,}0 \right)$$