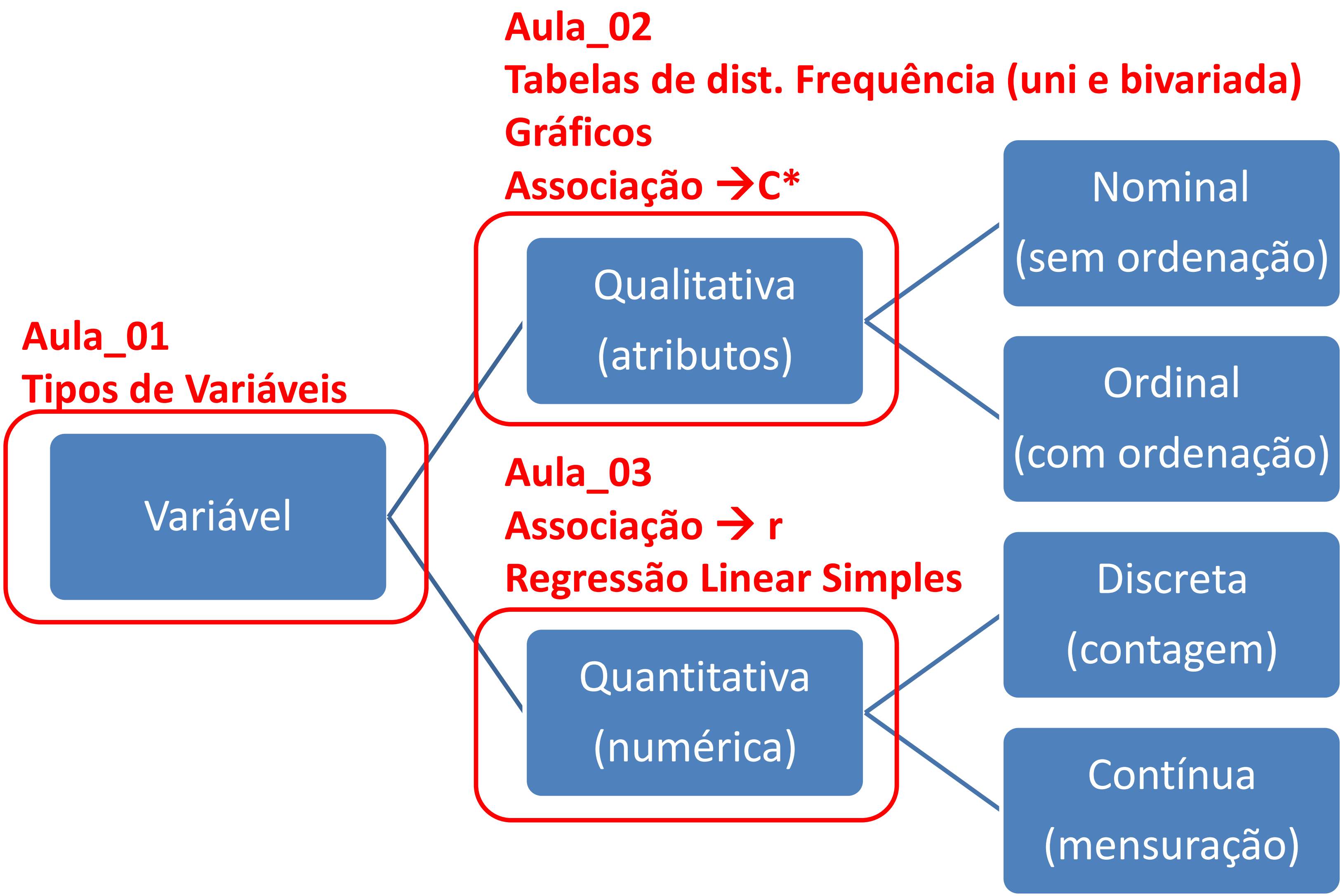
Escola Superior de Agricultura
"Luiz de Queiroz"
Universidade de São Paulo

LCE2112 – Estatística Aplicada às Ciências Sociais e Ambientais

Taciana Villela Savian
Sala 304, pav. Engenharia, ramal 237
tvsavian@usp.br
tacianavillela@gmail.com

Tipos de Variáveis



Variáveis Quantitativas Contínuas

1) Veremos nessa aula como resumir numericamente um conjunto de dados referentes à uma variável quantitativa contínua; (Aula_04)
2) Veremos como resumir, por meio de tabelas de classes de frequências, um conjunto de dados referentes à uma variável quantitativa contínua (Aula_05)
3) Como representar graficamente essas distribuições de frequências (Aula_05)

Variáveis Quantitativas Contínuas

Exemplo: Os resíduos industriais jogados nos rios, muitas vezes, absorvem o oxigênio necessário para respiração dos peixes e outras formas de vida aquática. Uma lei estadual exige um mínimo de 7 ppm de oxigênio dissolvido na água, a fim de que o conteúdo do mesmo seja suficiente para manter a vida aquática. Trinta amostras de água retiradas de um rio revelaram os seguintes quantidades:

Quadro1. Quantidade de oxigênio dissolvido (em ppm) de 30 amostras de água retiradas de um rio, em Piracicaba, em 2019.

4,5	6,8	7,8	8,1	5,9	8,8	7,7	9,2	5,9	6,8
8,3	8,1	7,9	7,1	6,9	5,8	7,5	7,5	7,5	6,9
8,1	7,9	4,8	5,7	6,9	6,8	7,4	7,3	7,3	7,2

Variáveis Quantitativas Contínuas

Quadro1. Quantidade de oxigênio dissolvido (em ppm) de 30 amostras de água retiradas de um rio, em Piracicaba, em 2019.

4,5	6,8	7,8	8,1	5,9	8,8	7,7	9,2	5,9	6,8
8,3	8,1	7,9	7,1	6,9	5,8	7,5	7,5	7,5	6,9
8,1	7,9	4,8	5,7	6,9	6,8	7,4	7,3	7,3	7,2

DADOS NÃO AGRUPADOS

- **Quantidade de informação muito grande e pouco clara;**
- **Resumir os dados sem perder informação:**
  - **Tabelas; (Tabela de Distribuição de Frequências – Dados Agrupados)**
  - **Números. (Medidas numéricas: medidas de posição/dispersão)**

Variáveis Quantitativas Contínuas

Quadro1. Quantidade de oxigênio dissolvido (em ppm) de 30 amostras de água retiradas de um rio, em Piracicaba, em 2019.

4,5	6,8	7,8	8,1	5,9	8,8	7,7	9,2	5,9	6,8
8,3	8,1	7,9	7,1	6,9	5,8	7,5	7,5	7,5	6,9
8,1	7,9	4,8	5,7	6,9	6,8	7,4	7,3	7,3	7,2

O primeiro passo é ordenar os dados (crescente) → dados elaborados (ou ROL)

Vamos usar um gráfico (diagrama de ramo-folhas) para ordenar os dados de forma mais eficiente....

Variáveis Quantitativas Contínuas

Um diagrama de ramos e folhas básico contém duas colunas separadas por uma linha vertical. A coluna da esquerda contém os ramos e a coluna da direita contém as folhas.

35, 78, 50, 63, 86, 73, 57, 82,
59, 75, 66, 79, 83, 71, 94, 59

Pode-se organizar este conjunto de dados utilizando uma representação gráfica do tipo seguinte:

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

Esta representação chama-se diagrama de caule-e-folhas.
O caule é a coluna com os números 3, 5, 6, 7, 8 e 9 que representam o algarismo das dezenas e as folhas que representam o algarismo das unidades de cada um dos dados.

Variáveis Quantitativas Contínuas

Quadro1. Quantidade de oxigênio dissolvido (em ppm) de 30 amostras de água retiradas de um rio, em Piracicaba, em 2019.

4,5	6,8	7,8	8,1	5,9	8,8	7,7	9,2	5,9	6,8
8,3	8,1	7,9	7,1	6,9	5,8	7,5	7,5	7,5	6,9
8,1	7,9	4,8	5,7	6,9	6,8	7,4	7,3	7,3	7,2

Ramo	Folhas
4	5, 8
5	7, 9, 8, 9
6	8, 9, 9, 8, 8, 9
7	9, 8, 9, 1, 7, 5, 4, 5, 3, 5, 3, 2
8	3, 1, 1, 1, 8
9	2

Ordenando as folhas

Ramo	Folhas
4	5, 8
5	7, 8, 9, 9
6	8, 8, 8, 9, 9, 9
7	1, 2, 3, 3, 4, 5, 5, 5, 7, 8, 9, 9
8	1, 1, 1, 3, 8
9	2

Variáveis Quantitativas Contínuas

DADOS BRUTOS

Quadro1. Quantidade de oxigênio dissolvido (em ppm) de 30 amostras de água retiradas de um rio, em Piracicaba, em 2019.

4,5	6,8	7,8	8,1	5,9	8,8	7,7	9,2	5,9	6,8
8,3	8,1	7,9	7,1	6,9	5,8	7,5	7,5	7,5	6,9
8,1	7,9	4,8	5,7	6,9	6,8	7,4	7,3	7,3	7,2

DADOS ELABORADOS (ROL)

Quadro1. Quantidade de oxigênio dissolvido (em ppm) de 30 amostras de água retiradas de um rio, em Piracicaba, em 2019.

4,5	4,8	5,7	5,8	5,9	5,9	6,8	6,8	6,8	6,9
6,9	6,9	7,1	7,2	7,3	7,3	7,4	7,5	7,5	7,5
7,7	7,8	7,9	7,9	8,1	8,1	8,1	8,3	8,8	9,2

Variáveis Quantitativas Contínuas

- **Média**

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

- **No exemplo:**

$$\bar{x} = \frac{\sum_{i=1}^{30} x_i}{30} = \frac{4,5 + 4,8 + 5,7 + \ldots + 9,2}{30} = \frac{214,4}{30}$$

$$\bar{x} = 7,1 \ \ ppm$$

Interpretação: Em média as 30 amostras de água apresentaram um teor de oxigênio dissolvido de 7,1 ppm.

OBS: É uma medida bastante afetada por valores extremos (muito alto ou muito baixo).

Variáveis Quantitativas Contínuas

- **Mediana**
  - **Observação que ocupa posição central na amostra de dados elaborados;**
  - **50% dos dados são menores ou iguais ao valor da mediana;**
  - **50% dos dados são maiores ou iguais ao valor da mediana;**
  - **Essa medida não é afetada por valores extremos já que depende apenas do tamanho da amostra e não dos valores da amostra em si;**

Variáveis Quantitativas Contínuas

- **Mediana**
  - **O primeiro passo é determinar a posição da mediana na amostra (depende do tamanho da amostra: par ou ímpar);**
  - **Posição da mediana (PMd):**

PMd=(n+1)/2

Amostra de tamanho ímpar: A= $\overbrace{\{2,2,5,7,7,11,100\}}^{n=7 \text{ elementos}}$

PMd = (n+1)/2 = (7+1)/2 = 4º valor da amostra ord.

Md = 7

Variáveis Quantitativas Contínuas

- **Mediana**
  - **O primeiro passo é determinar a posição da mediana na amostra (depende do tamanho da amostra: par ou ímpar);**
  - **Posição da mediana (PMd):**

    PMd=(n+1)/2

Amostra de tamanho par: $A=\overbrace{\{2,2,5,7,7,11\}}^{n=6\text{ elementos}}$

PMd = (n+1)/2 = (6+1)/2 = 3,5º valor da amostra ord. (existe?) A mediana é a média dos valores na 3ª e 4ª posição na amostra ordenada

Md = (5+7)/2 = 6.

Não necessariamente a Md está na amostra.

Variáveis Quantitativas Contínuas

- **Mediana**

DADOS ELABORADOS

Quadro1. Quantidade de oxigênio dissolvido (em ppm) de 30 amostras de água retiradas de um rio, em Piracicaba, em 2019.

4,5	4,8	5,7	5,8	5,9	5,9	6,8	6,8	6,8	6,9
6,9	6,9	7,1	7,2	7,3	7,3	7,4	7,5	7,5	7,5
7,7	7,8	7,9	7,9	8,1	8,1	8,1	8,3	8,8	9,2

Amostra de tamanho par (n=30)

Posição da Mediana: PMd = (30+1)/2 = 15,5º valor (existe?)

A Md é a média dos valores na 15ª e 16ª posição na amostra ordenada

Md = (7,3+7,3)/2 = 7,3 ppm

Variáveis Quantitativas Contínuas

- **Mediana**

DADOS ELABORADOS

Quadro1. Quantidade de oxigênio dissolvido (em ppm) de 30 amostras de água retiradas de um rio, em Piracicaba, em 2019.

4,5	4,8	5,7	5,8	5,9	5,9	6,8	6,8	6,8	6,9
6,9	6,9	7,1	7,2	7,3	7,3	7,4	7,5	7,5	7,5
7,7	7,8	7,9	7,9	8,1	8,1	8,1	8,3	8,8	9,2

Md = (7,3+7,3)/2 = 7,3 ppm

Interpretação: 50% das amostras de água apresentam teor de oxigênio dissolvido igual ou inferior a 7,3 ppm e os outros 50% apresentam esse teor igual ou superior a este valor (7,3 ppm).

Variáveis Quantitativas Contínuas

- **Mediana (Md) e os Quartis (Q)**

O pesquisador pode estar interessado em conhecer outros valores da amostra além desse valor central (mediana) que divide a amostra ao meio;

Valores extremos: menor e maior valor na amostra;

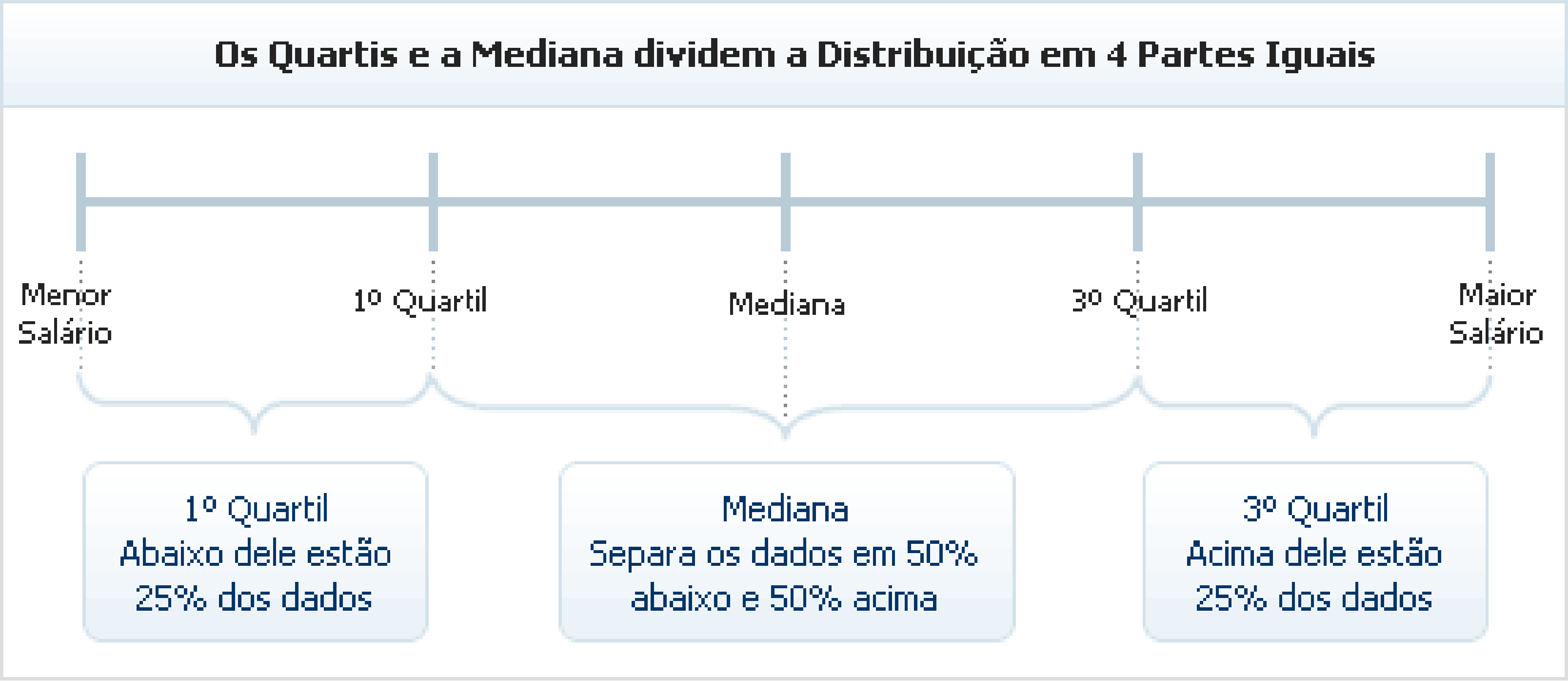
Primeiro Quartil (Q1): Valor que delimita os 25% menores valores na amostra;

Terceiro Quartil (Q3): Valor que delimita os 25% maiores valores na amostra;

A mediana também é conhecida como segundo quartil (Q2).

Variáveis Quantitativas Contínuas

- **Mediana (Md) e os Quartis (Q)**



Variáveis Quantitativas Contínuas

- **Quartis (Q)**

  Se a mediana divide a amostra em 50% acima e 50% abaixo, para calcular o primeiro e terceiro quartil basta calcular a mediana em cada sub-amostra

sub-amostra 1: valores abaixo da mediana

sub-amostra 2: valores acima da mediana

Variáveis Quantitativas Contínuas

- **Quartis (Q)**

A={2,2,5,7,7,11,11,12,13,15,18,20,22}

n=13

PMd = (13+1)/2 = 7º valor

Md = 11

Sub-amostra 1: A1={2,2,5,7,7,11,11} → Md=Q1

Sub-amostra 2: A2={11,12,13,15,18,20,22} → Md=Q3

Variáveis Quantitativas Contínuas

- **Quartis (Q)**

Sub-amostra 1: A1={2,2,5,7,7,11,11} → Md=Q1

n = 7

PMd = PQ1 = (n+1)/2 = (7+1)/2 = 4º valor da sub-amostra 1

Q1 = 7

Indica que 25% dos valores da amostra original (A) estão são iguais ou menores do que o valor 7.

Variáveis Quantitativas Contínuas

- **Quartis (Q)**

Sub-amostra 2: A2={11,12,13,15,18,20,22} → Md=Q3

n = 7

PMd = PQ3 = (n+1)/2 = (7+1)/2 = 4º valor da sub-amostra 2

Q3 = 15

Indica que 25% dos valores da amostra original (A) estão são iguais ou maiores do que o valor 15 (ou 75% dos valores são iguais ou inferiores ao valor 15).

Variáveis Quantitativas Contínuas

- **Primeiro Quartil (Q1)**

DADOS ELABORADOS

Quadro1. Quantidade de oxigênio dissolvido (em ppm) de 30 amostras de água retiradas de um rio, em Piracicaba, em 2019.

4,5	4,8	5,7	5,8	5,9	5,9	6,8	6,8	6,8	6,9
6,9	6,9	7,1	7,2	7,3	7,3	7,4	7,5	7,5	7,5
7,7	7,8	7,9	7,9	8,1	8,1	8,1	8,3	8,8	9,2

Sub-amostra 1: valores iguais ou inferior à mediana (Md=7,3)

n = 15

PQ1 = (15+1)/2 = 8º valor na sub-amostra 1

Q1 = 6,8 ppm

Interpretação: 25% das amostras de água apresentam teor de oxigênio dissolvido igual ou abaixo de 6,8 ppm.

Variáveis Quantitativas Contínuas

- **Terceiro Quartil (Q3)**

DADOS ELABORADOS

Quadro1. Quantidade de oxigênio dissolvido (em ppm) de 30 amostras de água retiradas de um rio, em Piracicaba, em 2019.

4,5	4,8	5,7	5,8	5,9	5,9	6,8	6,8	6,8	6,9
6,9	6,9	7,1	7,2	7,3	7,3	7,4	7,5	7,5	7,5
7,7	7,8	7,9	7,9	8,1	8,1	8,1	8,3	8,8	9,2

Sub-amostra 2: valores iguais ou superior à mediana (Md=7,3)

n = 15

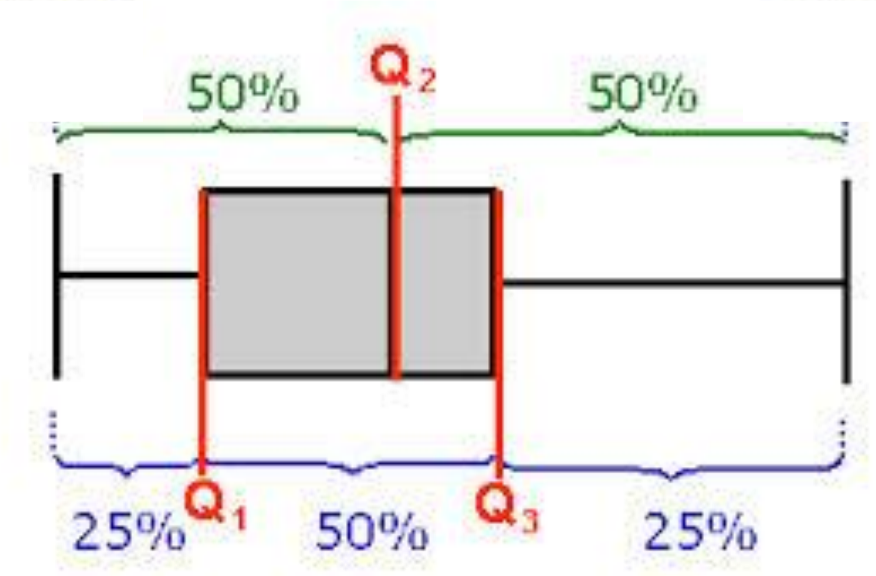
PQ3 = (15+1)/2 = 8º valor na sub-amostra 2

Q3 = 7,9 ppm

Interpretação: 25% das amostras de água apresentam teor de oxigênio dissolvido igual ou superior de 7,9 ppm.

Variáveis Quantitativas Contínuas

- **Boxplot:** **é um gráfico que possibilita representar a distribuição de um conjunto de dados (amostra) com base em algum de seus parâmetros descritivos: Mediana (Md=Q2), primeiro quartil (Q1), terceiro quartil (Q3) e intervalo interquartílico (IQ=Q3-Q1) bem como identificar se na amostra existe algum valor discrepante.**



Variáveis Quantitativas Contínuas

- **Boxplot**

Valores discrepantes são:

Valores abaixo de: Q1-1,5x(Q3-Q1)

Valores acima de: Q3+1,5x(Q3-Q1)

No exemplo:

Q1= 6,8 e Q3= 7,9

Valores abaixo de Q1-1,5x(Q3-Q1)

6,8-1,5x(7,9-6,8) = 5,2ppm (tem?)

Valores acima de Q3+1,5x(Q3-Q1)

7,9+1,5x(7,9-6,8) = 9,5 ppm (tem?)

Variáveis Quantitativas Contínuas

- **Boxplot**

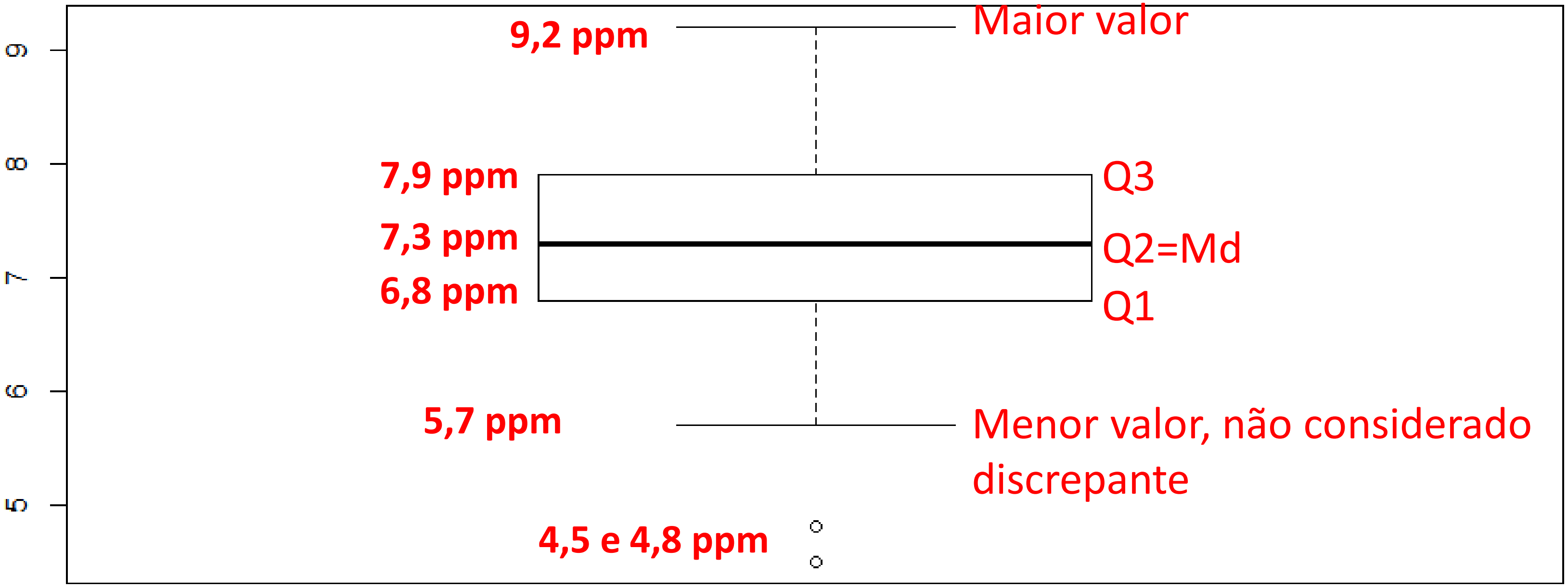


Figura 1. Boxplot para quantidade de oxigênio dissolvido (em ppm) de 30 amostras de água retiradas de um rio, em Piracicaba, em 2019.

Variáveis Quantitativas Contínuas

- **Moda (Mo)**

  É o valor que apresenta a maior frequência absoluta na amostra ordenada

Amostra unimodal = uma moda

Amostra bimodal = duas modas

Amostra multimodal = três ou mais modas

Amostra amodal = nenhum dos valores se repete.

Variáveis Quantitativas Contínuas

DADOS ELABORADOS (ROL)

Quadro1. Quantidade de oxigênio dissolvido (em ppm) de 30 amostras de água retiradas de um rio, em Piracicaba, em 2019.

4,5	4,8	5,7	5,8	5,9	5,9	**6,8**	**6,8**	**6,8**	**6,9**
6,9	**6,9**	7,1	7,2	7,3	7,3	7,4	**7,5**	**7,5**	**7,5**
7,7	7,8	7,9	7,9	**8,1**	**8,1**	**8,1**	8,3	8,8	9,2

Temos uma Amostra multimodal = três ou mais modas → Pois os valores 6,8; 6,9; 7,5 e 8,1 ppm se repetem três vezes cada um.

Variáveis Quantitativas Contínuas

- **Exercício**

Produção de resina, em kg, de 15 árvores de *Pinus elliottii*.

0,71	**2,63**	**3,63**	**1,94**	**3,69**
2,04	**3,94**	**4,05**	**1,80**	**2,06**
2,77	**1,42**	**1,20**	**2,48**	**3,77**

Calcule e interprete:

- *Média;*
- *Mediana;*
- *Primeiro quartil;*
- *Terceiro quartil;*
- *Verificar a existência de valores discrepantes;*

Variáveis Quantitativas Contínuas

- **Exercício**

Produção de resina, em kg, de 15 árvores de *Pinus elliottii*.

0,71	**2,63**	**3,63**	**1,94**	**3,69**
2,04	**3,94**	**4,05**	**1,80**	**2,06**
2,77	**1,42**	**1,20**	**2,48**	**3,77**

Calcule e interprete:

- *Média; (2,54kg)*
- *Mediana; (2,48kg)*
- *Primeiro quartil; (1,87kg)*
- *Terceiro quartil; (3,66kg)*
- *Verificar a existência de valores discrepantes; (Não há)*

Variáveis Quantitativas Contínuas

Quadro1. Densidade da madeira (g/cm³) de 30 clones de *Eucalyptus grandis* medidos em um plantio comercial, em Piracicaba, em 2015.

0,230	0,230	0,256	0,267	0,294	0,295	0,297	0,306	0,324	0,327
0,328	0,329	0,330	0,333	0,338	0,343	0,345	0,347	0,357	0,359
0,360	0,364	0,373	0,392	0,405	0,413	0,427	0,445	0,446	0,471

MEDIDAS DE DISPERSÃO (VARIABILIDADE)

- Amplitude (A);
- Variância (s^2);
- Desvio-padrão (s);
- Coeficiente de Variação (CV).

Grandezas numéricas que descrevem um conjunto de dados pela quantificação da sua variabilidade.

Variáveis Quantitativas Contínuas

- **Exemplo**

Elementos	Amostra 1	Amostra 2	Amostra 3
1	100	80	10
2	100	90	50
3	100	100	100
4	100	100	100
5	100	100	100
6	100	110	150
7	100	120	190
Média	100	100	100
Mediana	100	100	100
Moda	100	100	100

Variáveis Quantitativas Contínuas

Quadro1. Densidade da madeira (g/cm³) de 30 clones de *Eucalyptus grandis* medidos em um plantio comercial, em Piracicaba, em 2015.

0,230	0,230	0,256	0,267	0,294	0,295	0,297	0,306	0,324	0,327
0,328	0,329	0,330	0,333	0,338	0,343	0,345	0,347	0,357	0,359
0,360	0,364	0,373	0,392	0,405	0,413	0,427	0,445	0,446	0,471

MEDIDAS DE DISPERSÃO (VARIABILIDADE)

– **Amplitude (A); (Afetada por valores extremos)**

A = maior valor – menor valor

A= 0,471 – 0,230 = 0,241 g/cm³

Interpretação: O clone de *Eucalyptus grandis* que apresentou a maior densidade de madeira superou o de menor densidade em 0,241 g/cm³.

Variáveis Quantitativas Contínuas

Quadro1. Densidade da madeira (g/cm³) de 30 clones de *Eucalyptusgrandis*medidos em um plantio comercial, em Piracicaba, em 2015.

0,230	0,230	0,256	0,267	0,294	0,295	0,297	0,306	0,324	0,327
0,328	0,329	0,330	0,333	0,338	0,343	0,345	0,347	0,357	0,359
0,360	0,364	0,373	0,392	0,405	0,413	0,427	0,445	0,446	0,471

MEDIDAS DE DISPERSÃO (VARIABILIDADE)

– **Variância (s^2);** **(unidades dos dados ao quadrado)**

Medida de dispersão baseada nos desvios (d_i) das observações (x_i) em relação à media da amostra ($\bar{x}$).

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

Variáveis Quantitativas Contínuas

Quadro1. Densidade da madeira (g/cm³) de 30 clones de *Eucalyptusgrandis*medidos em um plantio comercial, em Piracicaba, em 2015.

0,230	**0,230**	**0,256**	**0,267**	**0,294**	**0,295**	**0,297**	**0,306**	**0,324**	**0,327**
0,328	**0,329**	**0,330**	**0,333**	**0,338**	**0,343**	**0,345**	**0,347**	**0,357**	**0,359**
0,360	**0,364**	**0,373**	**0,392**	**0,405**	**0,413**	**0,427**	**0,445**	**0,446**	**0,471**

– **Variância (s^2); (unidades dos dados ao quadrado)**

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

$$\sum_{i=1}^{n} {x_i}^2 = {x_1}^2 + {x_2}^2 + \cdots . + {x_n}^2 = 0{,}230^2 + \cdots + 0{,}471^2 = 3{,}6637$$

$$\left(\sum_{i=1}^{n} x_i\right)^2 = (x_1 + x_2 + \cdots + x_n)^2 = (0{,}230 + \cdots + 0{,}471)^2 = 106{,}7296$$

Variáveis Quantitativas Contínuas

Quadro1. Densidade da madeira (g/cm³) de 30 clones de *Eucalyptusgrandis*medidos em um plantio comercial, em Piracicaba, em 2015.

0,230	0,230	0,256	0,267	0,294	0,295	0,297	0,306	0,324	0,327
0,328	0,329	0,330	0,333	0,338	0,343	0,345	0,347	0,357	0,359
0,360	0,364	0,373	0,392	0,405	0,413	0,427	0,445	0,446	0,471

– **Variância (s^2); (unidades dos dados ao quadrado)**

$$s^2 = \frac{\sum_{i=1}^{n} {x_i}^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{3{,}6637 - \frac{106{,}7296}{30}}{29}$$

$$= 0{,}0037\ (g/cm^3)^2$$

Interpretação: é uma medida de variabilidade ou dispersão mas de difícil interpretação pois a unidade está ao quadrado!

Variáveis Quantitativas Contínuas

Quadro1. Densidade da madeira (g/cm³) de 30 clones de *Eucalyptusgrandis*medidos em um plantio comercial, em Piracicaba, em 2015.

0,230	0,230	0,256	0,267	0,294	0,295	0,297	0,306	0,324	0,327
0,328	0,329	0,330	0,333	0,338	0,343	0,345	0,347	0,357	0,359
0,360	0,364	0,373	0,392	0,405	0,413	0,427	0,445	0,446	0,471

– **Desvio padrão (s);**

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

$$s = \sqrt{0{,}0037\ (g/cm^3)^2} = 0{,}0605 g/cm^3$$

Variáveis Quantitativas Contínuas

Quadro1. Densidade da madeira (g/cm³) de 30 clones de *Eucalyptusgrandis*medidos em um plantio comercial, em Piracicaba, em 2015.

0,230	**0,230**	**0,256**	**0,267**	**0,294**	**0,295**	**0,297**	**0,306**	**0,324**	**0,327**
0,328	**0,329**	**0,330**	**0,333**	**0,338**	**0,343**	**0,345**	**0,347**	**0,357**	**0,359**
0,360	**0,364**	**0,373**	**0,392**	**0,405**	**0,413**	**0,427**	**0,445**	**0,446**	**0,471**

– **Desvio padrão (s);** $\boldsymbol{s = 0,0605 g/cm^3}$

Quando o desvio padrão é pequeno, próximo de zero, existirá uma grande concentração dos dados em torno da média e se o desvio padrão for grande os valores não se concentrarão com tal intensidade em torno da média.

Variáveis Quantitativas Contínuas

- **Coeficiente de variação (CV)**
- O desvio padrão e a variância são medidas de variabilidade absoluta dos dados, ou seja, são medidas dependentes da grandeza, escala ou unidade de medida dos dados;
- Conjunto de dados com diferentes unidades de medida não podem ter suas medidas de dispersão comparadas pela variância ou pelo desvio padrão;
- É necessário uma medida de dispersão que seja relativa, ou seja, que não dependa da unidade dos dados.

Variáveis Quantitativas Contínuas

- **Coeficiente de variação (CV)**
- O coeficiente de variação (CV) é utilizado para esse propósito e é dado por:

$$CV = \frac{S}{\bar{x}} * 100$$

Estatística Descritiva

- **Coeficiente de variação (CV)**

Exemplo 1: Foram medidos o peso, em kg, de 10 pessoas formando duas amostras.

Amostra A: n= 5 pessoas

Peso médio= 70kg

Desvio padrão = 3kg

Qual amostra apresenta maior variabilidade?

Amostra B: n= 5 pessoas

Peso médio = 70kg

Desvio padrão = 5kg

Estatística Descritiva

- **Coeficiente de variação ($\boldsymbol{CV}$)**

Exemplo 2: Foram medidos o peso, em kg, de 10 pessoas formando duas amostras.

Amostra C: n= 5 pessoas

Peso médio= 70kg

Desvio padrão = 3kg

$$\boldsymbol{CV = {}^{3}/_{70} \times 100 = 4,28\%}$$

Qual amostra apresenta maior variabilidade?

Amostra D: n= 5 pessoas

Peso médio = 86kg

Desvio padrão = 2kg

$$\boldsymbol{CV = {}^{2}/_{86} \times 100 = 2,33\%}$$

Estatística Descritiva

- **Coeficiente de variação (CV)**

Exemplo 3: Foram medidos o peso, em kg, e altura, em cm, de 5 pessoas formando duas amostras.

Amostra E: n= 5 pessoas

Peso médio= 70kg

Desvio padrão = 3kg

$$\boldsymbol{CV = {}^{3}/_{70} \times 100 = 4,28\%}$$

Qual amostra apresenta maior variabilidade?

Amostra F: n= 5 pessoas

Altura média = 168cm

Desvio padrão = 5cm

$$\boldsymbol{CV = {}^{5}/_{168} \times 100 = 2,98\%}$$

Estatística Descritiva

- **Coeficiente de variação (CV)**
- **No exemplo densidade de madeira:**

$$\bar{x} = 0,344 \frac{g}{cm^3}$$

$$s = 0,0605 \frac{g}{cm^3}$$

$$CV = {}^{0,0605}/_{0,344} \times 100 = 17,59\%$$

Variáveis Quantitativas Contínuas

- **Exercício**

Produção de resina, em kg, de 15 árvores de *Pinus elliottii*.

0,71	**2,63**	**3,63**	**1,94**	**3,69**
2,04	**3,94**	**4,05**	**1,80**	**2,06**
2,77	**1,42**	**1,20**	**2,48**	**3,77**

Calcular e interpretar as medidas de variabilidade:

- *Amplitude;*
- *Variância;*
- *Desvio padrão;*
- *Coeficiente de Variação.*

Variáveis Quantitativas Contínuas

- **Exercício**

Produção de resina, em kg, de 15 árvores de *Pinus elliottii*.

0,71	**2,63**	**3,63**	**1,94**	**3,69**
2,04	**3,94**	**4,05**	**1,80**	**2,06**
2,77	**1,42**	**1,20**	**2,48**	**3,77**

Calcular e interpretar as medidas de variabilidade:

- *Amplitude* ***(A=3,34 kg)****;*
- *Variância* ***(*** $s^2 = 1,15 kg^2$ ***)****;*
- *Desvio padrão* ***(s=1,07kg)****;*
- *Coeficiente de Variação* ***(CV=42,20%)****.*

Estatística Descritiva

- **Nessa Aula (Aula 05) veremos como construir uma Tabela de Distribuição de Classes de Frequências para uma variável quantitativa contínua.**

  Feita a coleta de dados, os mesmos apresentam-se de maneira desorganizada (sem um padrão crescente/decrescente), ainda sem valor informativo sobre o fenômeno em estudo.

  O **primeiro passo** para construção da tabela de distribuição de frequências é a ordenação dos dados (o mais usual é em ordem crescente).

Variáveis Quantitativas Contínuas

Exemplo: Os resíduos industriais jogados nos rios, muitas vezes, absorvem o oxigênio necessário para respiração dos peixes e outras formas de vida aquática. Uma lei estadual exige um mínimo de 7 ppm de oxigênio dissolvido na água, a fim de que o conteúdo do mesmo seja suficiente para manter a vida aquática. Trinta amostras de água retiradas de um rio revelaram os seguintes quantidades:

Quadro1. Quantidade de oxigênio dissolvido (em ppm) de 30 amostras de água retiradas de um rio, em Piracicaba, em 2019. → **ROL**

4,5	4,8	5,7	5,8	5,9	5,9	6,8	6,8	6,8	6,9
6,9	6,9	7,1	7,2	7,3	7,3	7,4	7,5	7,5	7,5
7,7	7,8	7,9	7,9	8,1	8,1	8,1	8,3	8,8	9,2

Estatística Descritiva

- **1º passo: Ordenar os valores da amostra (ROL)**
- **2º passo: Determinar o número de classes (k)**

Decidir quantas classes (k) terá a tabela de distribuição de frequências. Embora essa decisão seja arbitrária (o pesquisador pode escolher). Existem algumas fórmulas para auxiliar, por exemplo:

$$k = \sqrt{n}$$ **em que *n* é o tamanho da amostra**

ou

$$k = 1 + 3,32\log_{10}(n)$$ **Fórmula de Sturges**

<u>**Usaremos a Fórmula de Sturges!**</u>

Estatística Descritiva

- **2º passo: Determinar o número de classes (k)**

<u>Usaremos a Fórmula de Sturges!</u>

$$k = 1 + 3,32\log_{10}(n)$$

$$k = 1 + 3,32\log_{10}(30)$$

$$k = 5,9 \ \ classes \Rightarrow 6 \ \ classes$$

Classes	Freq.
Classe 1	f1
Classe 2	f2
Classe 3	f3
Classe 4	f4
Classe 5	f5
Classe 6	f6
Total	f.

Cada uma dessas classes representa um intervalo de valores da variável, precisamos definir a amplitude (tamanho) de cada intervalo, ou seja, o tamanho de cada classe.

Estatística Descritiva

- **3º passo: Amplitude de classe (h)**

Tabela 1. Distribuição de Classes de Frequências

Classes	Freq.
Classe 1	f1
Classe 2	f2
Classe 3	f3
Classe 4	f4
Classe 5	f5
Classe 6	f6
Total	f.

Qual o tamanho de cada classe = Qual a amplitude de cada classe (*h*)?

$$h = \frac{A}{k}$$

"A" representa a amplitude dos dados

A = Maior valor – Menor valor

A= 9,2 – 4,5 = 4,7

e...**k (é número de classes) =6**

$$h = \frac{A}{k} = \frac{4,7}{6} = 0,8$$

Ou seja, cada uma das seis classes terá uma amplitude de 0,8 unidades.

Estatística Descritiva

- **4º passo: Construção da 1ª classe**

Tabela 1. Distribuição de frequência.

Classes	Freq.
Classe 1	f1
Classe 2	f2
Classe 3	f3
Classe 4	f4
Classe 5	f5
Classe 6	f6
Total	f.

Primeira Classe: o menor valor dos dados (x_{min}=4,5) deverá ser o limite inferior da primeira classe.

Classe 1: [x_{min} ; x_{min} + h)

Lembrando: h é a amplitude de classe (3º passo → h=0,8)

Desta forma, a classe 1 é representada pelo seguinte intervalo:

Classe 1: [4,5 ; 4,5 + 0,8)
[4,5 ; 5,3)

Estatística Descritiva

- **4º passo: Construção da 1ª classe**

Estabelecer uma notação para representação das classes na tabela.

- Parêntesis: indica que o valor não pertence à classe em questão
- Colchetes: indica que o valor pertence ao intervalo.

(1 ; 3) é a notação de uma classe em que nem o valor 1, nem o valor 3 pertencem ao intervalo;

(1 ; 3] é a notação de uma classe em que o valor 1 não pertence ao intervalo mas o valor 3 pertence;

[1 ; 3) é a notação de uma classe em que o valor 1 pertence ao intervalo mas o valor 3 não pertence; (usaremos essa)

[1 ; 3] é a notação de uma classe em que ambos os valores, 1 e 3, pertencem ao intervalo;

Estatística Descritiva

- **Quarto passo: Construção da 1ª classe**

Tabela 1. Distribuição de frequência.

Classes	Freq.
[4,5 ; 5,3)	f1
Classe 2	f2
Classe 3	f3
Classe 4	f4
Classe 5	f5
Classe 6	f6
Total	f.

Primeira Classe: o menor valor dos dados (x_{min}=4,5) deverá ser o limite inferior da primeira classe.

Classe 1: [x_{min} ; x_{min} + h)

Lembrando: h é a amplitude de classe (3º passo → h=0,8)

Desta forma, a classe 1 é representada pelo seguinte intervalo:

Classe 1: [4,5 ; 4,5 + 0,8)
[4,5 ; 5,3)

Estatística Descritiva

- **5º passo: Construção das demais classes**

Tabela 1. Distribuição de frequência.

Classes	Freq.
[4,5 ; 5,3)	f1
[5,3 ; 6,1)	f2
[6,1 ; 6,9)	f3
[6,9 ; 7,7)	f4
[7,7 ; 8,5)	f5
[8,5 ; 9,3)	f6
Total	f.

Para construção das demais (k = 6) classes da tabela basta transpor o limite superior de cada classe para o limite inferior da classe subsequente e somar a esse valor a amplitude de classe (h = 0,8).

1ª classe: (4,5 ; 5,3]
2ª classe: (5,3 ; 6,1]

5,3 + 0,8 = 6,1

Estatística Descritiva

- **6º passo: Determinar as frequências em cada classe (contar quantos valores tem em cada classe construída)**

Tabela 1. Distribuição de frequência.

Classes	Freq.
[4,5 ; 5,3)	f1 = 2
[5,3 ; 6,1)	f2
[6,1 ; 6,9)	f3
[6,9 ; 7,7)	f4
[7,7 ; 8,5)	f5
[8,5 ; 9,3)	f6
Total	f.

Quadro1. Quantidade de oxigênio dissolvido (em ppm) de 30 amostras de água retiradas de um rio, em Piracicaba, em 2019. → **ROL**

4,5	6,9	7,7
4,8	6,9	7,8
5,7	7,1	7,9
5,8	7,2	7,9
5,9	7,3	8,1
5,9	7,3	8,1
6,8	7,4	8,1
6,8	7,5	8,3
6,8	7,5	8,8
6,9	7,5	9,2

Estatística Descritiva

- **6º passo: Determinar as frequências em cada classe (contar quantos valores tem em cada classe construída)**

Tabela 1. Distribuição de frequência.

Classes	Freq.
[4,5 ; 5,3)	f1 = 2
[5,3 ; 6,1)	f2 = 4
[6,1 ; 6,9)	f3
[6,9 ; 7,7)	f4
[7,7 ; 8,5)	f5
[8,5 ; 9,3)	f6
Total	f.

Quadro1. Quantidade de oxigênio dissolvido (em ppm) de 30 amostras de água retiradas de um rio, em Piracicaba, em 2019. → ROL

4,5	6,9	7,7
4,8	6,9	7,8
5,7	7,1	7,9
5,8	7,2	7,9
5,9	7,3	8,1
5,9	7,3	8,1
6,8	7,4	8,1
6,8	7,5	8,3
6,8	7,5	8,8
6,9	7,5	9,2

Estatística Descritiva

- **6º passo: Determinar as frequências em cada classe (contar quantos valores tem em cada classe construída)**

Tabela 1. Distribuição de frequência.

Classes	Freq.
[4,5 ; 5,3)	f1 = 2
[5,3 ; 6,1)	f2 = 4
[6,1 ; 6,9)	f3 = 3
[6,9 ; 7,7)	f4
[7,7 ; 8,5)	f5
[8,5 ; 9,3)	f6
Total	f.

Quadro1. Quantidade de oxigênio dissolvido (em ppm) de 30 amostras de água retiradas de um rio, em Piracicaba, em 2019. → ROL

4,5	6,9	7,7
4,8	6,9	7,8
5,7	7,1	7,9
5,8	7,2	7,9
5,9	7,3	8,1
5,9	7,3	8,1
6,8	7,4	8,1
6,8	7,5	8,3
6,8	7,5	8,8
6,9	7,5	9,2

Estatística Descritiva

- **6º passo: Determinar as frequências em cada classe (contar quantos valores tem em cada classe construída)**

Tabela 1. Distribuição de frequência.

Classes	Freq.
[4,5 ; 5,3)	f1 = 2
[5,3 ; 6,1)	f2 = 4
[6,1 ; 6,9)	f3 = 3
[6,9 ; 7,7)	f4 = 11
[7,7 ; 8,5)	f5
[8,5 ; 9,3)	f6
Total	f.

Quadro1. Quantidade de oxigênio dissolvido (em ppm) de 30 amostras de água retiradas de um rio, em Piracicaba, em 2019. → ROL

4,5	6,9	7,7
4,8	6,9	7,8
5,7	7,1	7,9
5,8	7,2	7,9
5,9	7,3	8,1
5,9	7,3	8,1
6,8	7,4	8,1
6,8	7,5	8,3
6,8	7,5	8,8
6,9	7,5	9,2

Estatística Descritiva

- **6º passo: Determinar as frequências em cada classe (contar quantos valores tem em cada classe construída)**

Tabela 1. Distribuição de frequência.

Classes	Freq.
[4,5 ; 5,3)	f1 = 2
[5,3 ; 6,1)	f2 = 4
[6,1 ; 6,9)	f3 = 3
[6,9 ; 7,7)	f4 = 11
[7,7 ; 8,5)	f5 = 8
[8,5 ; 9,3)	f6
Total	f.

Quadro1. Quantidade de oxigênio dissolvido (em ppm) de 30 amostras de água retiradas de um rio, em Piracicaba, em 2019. → **ROL**

4,5	6,9	7,7
4,8	6,9	7,8
5,7	7,1	7,9
5,8	7,2	7,9
5,9	7,3	8,1
5,9	7,3	8,1
6,8	7,4	8,1
6,8	7,5	8,3
6,8	7,5	8,8
6,9	7,5	9,2

Estatística Descritiva

- **6º passo: Determinar as frequências em cada classe (contar quantos valores tem em cada classe construída)**

Tabela 1. Distribuição de frequência.

Classes	Freq.
[4,5 ; 5,3)	f1 = 2
[5,3 ; 6,1)	f2 = 4
[6,1 ; 6,9)	f3 = 3
[6,9 ; 7,7)	f4 = 11
[7,7 ; 8,5)	f5 = 8
[8,5 ; 9,3)	f6 = 2
Total	f.

Quadro1. Quantidade de oxigênio dissolvido (em ppm) de 30 amostras de água retiradas de um rio, em Piracicaba, em 2019. → **ROL**

4,5	6,9	7,7
4,8	6,9	7,8
5,7	7,1	7,9
5,8	7,2	7,9
5,9	7,3	8,1
5,9	7,3	8,1
6,8	7,4	8,1
6,8	7,5	8,3
6,8	7,5	8,8
6,9	7,5	9,2

Estatística Descritiva

- **6º passo: Determinar as frequências em cada classe (contar quantos valores tem em cada classe construída)**

Tabela 1. Distribuição de frequência.

Classes	Freq.
[4,5 ; 5,3)	f1 = 2
[5,3 ; 6,1)	f2 = 4
[6,1 ; 6,9)	f3 = 3
[6,9 ; 7,7)	f4 = 11
[7,7 ; 8,5)	f5 = 8
[8,5 ; 9,3)	f6 = 2
Total	f. = 30

Quadro1. Quantidade de oxigênio dissolvido (em ppm) de 30 amostras de água retiradas de um rio, em Piracicaba, em 2019. → **ROL**

4,5	6,9	7,7
4,8	6,9	7,8
5,7	7,1	7,9
5,8	7,2	7,9
5,9	7,3	8,1
5,9	7,3	8,1
6,8	7,4	8,1
6,8	7,5	8,3
6,8	7,5	8,8
6,9	7,5	9,2

Estatística Descritiva

- **Completando a tabela com os demais tipos de frequências (relativa e percentual)**

Tabela 1. Distribuição de frequência da quantidade de oxigênio dissolvido (em ppm) de 30 amostras de água retiradas de um rio, em Piracicaba, em 2019

Classes	Frequência absoluta (f_i)	Frequência relativa (fr_i)	Frequência percentual ($f\%_i$)
[4,5 ; 5,3)	**2**	**2/30 = 0,07**	**0,07 x 100 = 7%**
[5,3 ; 6,1)	**4**	**4/30 = 0,13**	**0,13 x 100 = 13%**
[6,1 ; 6,9)	**3**	**3/30 = 0,10**	**0,10 x 100 = 10%**
[6,9 ; 7,7)	**11**	**11/30 = 0,37**	**0,37 x 100 = 37%**
[7,7 ; 8,5)	**8**	**8/30 = 0,27**	**0,27 x 100 = 27%**
[8,5 ; 9,3)	**2**	**2/30 = 0,07**	**0,07 x 100 = 7%**
Total	**30**	**1,01 ≈ 1,00**	**≈ 100%**

Que tipo de gráfico usar?

- Gráficos: Visualização mais sugestiva do que tabelas;
- Forma alternativa de apresentar as distribuições de frequências;
- Tipo de gráfico também irá depender do tipo e número de variáveis em estudo.
  - Qualitativas: gráfico de setores/barras/colunas (Já vimos em aulas anteriores);
  - Quantitativas: Histogramas, Polígonos de frequência, Ogivas (para frequências acumuladas).

Que tipo de gráfico usar?

- **Histogramas:** é uma sequência de retângulos postos lado a lado em que cada retângulo tem como base a amplitude da classe (h=0,8) e como altura a frequência (absoluta ou relativa);

Que tipo de gráfico usar?

- **HISTOGRAMA**

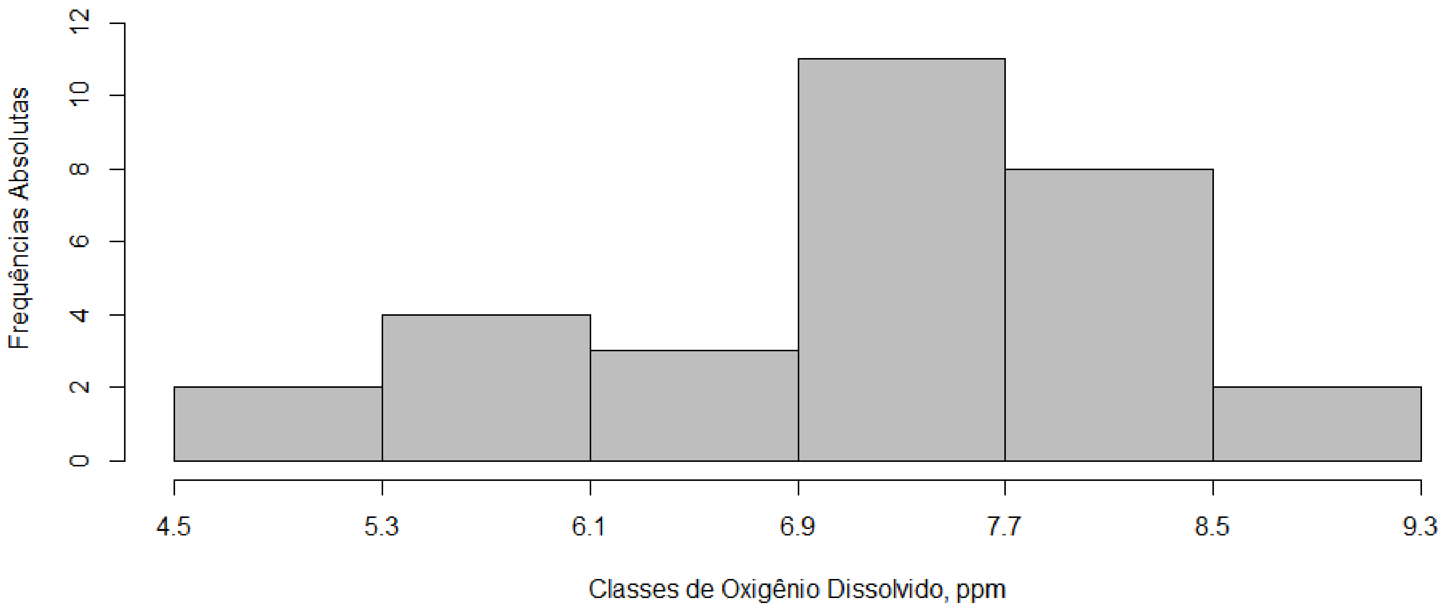


Figura 1. Histograma distribuição de frequência da quantidade de oxigênio dissolvido (em ppm) de 30 amostras de água retiradas de um rio, em Piracicaba, em 2019.

Exercício_Aula

- **Exercício**

Produção de resina, em kg, de 15 árvores de *Pinus elliottii*.

0,71	**2,63**	**3,63**	**1,94**	**3,69**
2,04	**3,94**	**4,05**	**1,80**	**2,06**
2,77	**1,42**	**1,20**	**2,48**	**3,77**

i) Construir a tabela de distribuição de frequências para os dados de densidade da madeira (Absoluta, Relativa e Percentual);

ii) Faça um histograma para representar a distribuição de frequência.

Exercício_via e-disciplina (até 21/09)

Quadro1. Densidade da madeira (g/cm³) de 40 clones de *Pinus* medidos em um plantio comercial, em Piracicaba, em 2015.

0,241	**0,174**	**0,199**	**0,265**	**0,239**	**0,258**	**0,244**	**0,246**	**0,293**	**0,231**
0,305	**0,265**	**0,195**	**0,368**	**0,213**	**0,395**	**0,407**	**0,303**	**0,224**	**0,131**
0,345	**0,228**	**0,260**	**0,345**	**0,158**	**0,229**	**0,370**	**0,407**	**0,429**	**0,331**
0,445	**0,427**	**0,559**	**0,426**	**0,356**	**0,229**	**0,391**	**0,376**	**0,348**	**0,290**

i) Calcule e interprete as medidas de tendência central e dispersão;

ii) Verifique a existência de valores discrepantes;

iii) Construa a tabela de distribuição de frequências para os dados de densidade da madeira (Absoluta, Relativa e Percentual);

iv) Faça um histograma para representar a distribuição de frequência.