

Profa. Clayde Regina Mendes

ESTATÍSTICA APLICADA
AULA 2

ORGANIZAÇÃO DE DADOS E MEDIDAS ESTATÍSTICAS

Considere a seguinte situação de pesquisa: Foram coletados os dados de um grupo de estudantes de uma faculdade e, após a coleta, os dados foram dispostos na seguinte BASE DE DADOS:

Número	Idade (anos)	Peso (kg)	Altura (cm)	Turno em que estuda	Cidade em que reside
1	19	50	166	2	6
2	17	85	183	1	7
3	20	83	174	1	9
4	18	52	163	3	9
5	17	60	164	1	7
6	17	55	170	1	8
7	17	62		3	1
8	19	85	173	1	9
9	18	53	161	1	3
10	18	49	160	2	8
11	18	48	160	1	9
12	17	52	160	2	7
13	19	48	161	1	7
14	18	51	163	2	2
15	19	60	182	2	9
16	20	62	167	1	9
17	19	55	169	3	8
18	18	90	176	1	7
19	22	58	167	3	5
20	18	52	172	2	7
21	18	56	165	3	7
22	19	53	168	2	5
23	18	52	166	1	9
24	18	55	166	1	7
25	18	68	175	1	9
26	20	46	162	1	9
27	17	47	158	1	7
28	17	51	162	1	9
29	18	45	161	2	9
30	17	58	160	2	5
31	18	58	160	3	1
32	20	47	160	1	9
33	18	47	155	3	3
34	18	58	163	1	1
35	20	63	169	1	9
36	18	60	170	3	9
37	24	53	163	3	7
38	18	69	160	1	9

39	19	57	160	1	3
40	21	76	163	2	7
41	25	63	163	1	8
42	17	49	162	3	1
43	17	54	177	1	9
44	18	60	170	3	9
45	19	63	170	1	9
46	18		170	1	9
47	19	54	164	1	9
48	19	50	161	3	9
49	19	70	187	1	9
50	17	60	174	3	9
51	19	65	168	1	9
52	18	60	173	1	9
53	19	80	181	3	9
54	20	79	180	1	9
55	19	45	163	2	9
56	20	77	177	1	9
57	17	52	170	1	9
58	18	75	180	2	9
59	18	56	170	3	2
60	19	52	178	1	9
61	19	59	164	3	8
62	17	58	176	1	1
63	19	50	162	2	9
64	19	50	167	2	9
65	21	57	160	1	8
66	18	49	158	2	7
67	18	65	177	2	9
68	17	53	162	2	9
69	20	58	163	3	9
70	19	86	175	1	9

Categorização/codificação:

Turno em que estuda	Cidade de Residência
1 – matutino	1 – Hortolândia
	2 – Sumaré
	3 – Valinhos
	4 – Louveira
2 – noturno	5 – Jundiaí
	6 – Itatiba
	7 – Salto
3 – vespertino	8 – Indaiatuba
	9 – Campinas

Essa categorização numérica facilita inclusive a transferência de uma base de dados organizada no Excel para um programa estatístico, por exemplo, para o SPSS.

Resumindo: Observando essa base de dados, quais foram as variáveis coletadas e de que tipo elas são?

Nessa base de dados, considere apenas a coluna referente aos dados da variável idade.

Observe que esses valores numéricos não estão dispostos de forma ordenada.

Alguns autores utilizam a expressão **numericamente desorganizados** para dados numéricos que não seguem uma ordem nem crescente, nem decrescente.

Em Estatística esses dados desorganizados são chamados de **dados brutos**.

Como é muito difícil trabalhar estatisticamente com dados desorganizados, a primeira forma de organização é ordenar (de forma crescente ou decrescente) os dados brutos. A forma mais comum de organização é em ordem crescente.

Os dados organizados recebem o nome de **ROL**

Exercício: Construa o rol dos dados coletados para a variável idade (utilize ordem crescente)

Observando o rol que você acabou de construir, responda:

1. Quais são os valores (diferentes) da variável idade?
2. Quantos valores (diferentes) a variável idade apresentou?

Para um número “pequeno” de valores diferentes de uma variável quantitativa, a melhor forma de organização é através de uma **tabela ponto-a-ponto**.

Exercício: Organize os dados da variável idade em uma tabela. Não esqueça que ela deve ser devidamente formatada.

Tabela 1 - Distribuição da idade dos estudantes

Um lembrete: apesar de existirem três formas para se expressar a frequência relativa, em trabalhos de pesquisa a forma mais usada é em porcentagem.

Uma tabela desse tipo em que a variável apresenta exatamente os valores obtidos na coleta de dados acompanhados de suas respectivas frequências é chamada uma **tabela ponto a ponto** ou **discreta**.

Para representarmos graficamente os dados de uma tabela ponto a ponto, o melhor gráfico é o de **colunas simples**.

EXERCÍCIO: CONSTRUA UM GRÁFICO QUE REPRESENTA OS DADOS DA TABELA 1 (relativa às idades dos alunos de uma faculdade)

PROBLEMA: Como calcular a média das idades que estão agrupadas na Tabela 1?

Para responder a essa questão, é preciso pensar como seria o cálculo da média se, ao invés de utilizarmos os dados agrupados na Tabela 1, utilizássemos o rol.

$$\bar{X} = \frac{\sum f_i x_i}{N}, \text{ lembrando que:}$$

x_i significa cada um dos valores da variável,

f_i significa a frequência de cada um desses valores da variável,

N é a frequência total.

(O símbolo \bar{X} representa a palavra **Média**)

Exemplo: Considere a tabela referente às idades. Para calcular a idade média desses estudantes vamos trabalhar com uma tabela auxiliar.

Idade (anos)	Número de alunos	
17	14	
18	24	
19	19	
20	8	
21	2	
22	1	
24	1	
25	1	
TOTAL	70	

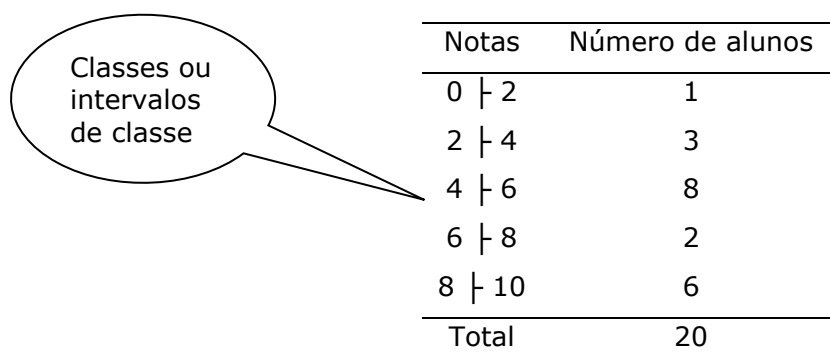
Agora vamos passar para a variável altura.

• QUANTOS VALORES DIFERENTES A VARIÁVEL ALTURA APRESENTA?

O problema das tabelas ponto a ponto ocorre quando precisamos organizar e resumir um “grande” número de valores distintos de uma variável quantitativa, como, por exemplo, da variável altura, que foi coletada com os alunos dessa mesma Faculdade.

Para esses casos, recorreremos à organização dos dados em **tabelas em classes**.

Antes de organizarmos os dados coletados das estaturas em uma tabela em classes, vamos apresentar um exemplo de uma tabela desse tipo:

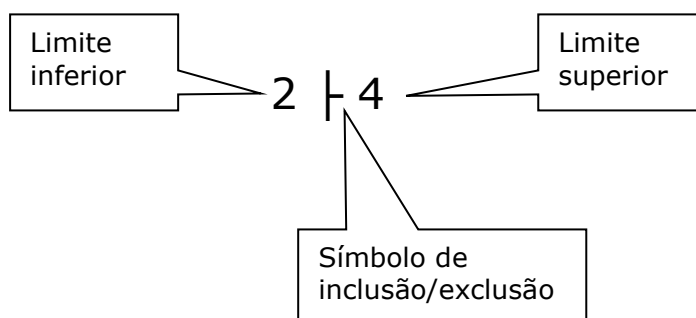


Classes ou intervalos de classe

Notas	Número de alunos
0 2	1
2 4	3
4 6	8
6 8	2
8 10	6
Total	20

Cada classe é formada por dois limites numéricos unidos por um símbolo de intervalo. Utilizaremos o símbolo | sugerido pelo IBGE, para relacionar os limites inferior e superior da classe. Ele é chamado um símbolo de inclusão/exclusão, por incluir na classe o limite inferior e excluir da mesma o limite superior.

Exemplo: Consideremos a classe:



IMPORTANTE:

(1) A cada classe corresponde um **ponto médio**, geralmente indicado por x_i e calculado como:

$$x_i = \frac{\text{limite inferior} + \text{limite superior}}{2};$$

(2) A cada classe corresponde uma **amplitude de classe**, muitas vezes indicada por **h**, que é calculada como: **h = lim sup - lim inf**

Exercício: Calcule o ponto médio e a amplitude de classe para 4 | 6.

Agora podemos começar a organizar em classes, os dados referentes às alturas, que constam da sua BASE DE DADOS do início deste capítulo.

1º passo: determinar o número de classes a ser utilizado na tabela. Podemos usar a regra de Sturges que é dada por: $i = 1 + 3,3 \cdot \log n$, na qual

i = número de classes

n = número total de dados

Em relação às alturas, determine:

Número de dados coletados:

Número de classes:

2º passo: determinação da amplitude do intervalo de classe

Para determinarmos **h** a ser utilizado, precisamos dos valores **máximo** e **mínimo** observados nos dados brutos (ou seja, nos dados coletados), pois a determinação da amplitude de classe, em geral, é dada

por:
$$h = \frac{\text{valor máximo} - \text{valor mínimo}}{i}$$

Determine:

Maior valor da altura = (esse é o valor máximo coletado para as alturas)

Menor valor da altura = (esse é o valor mínimo coletado para alturas)

Daí, **h** =

Caso o resultado dessa divisão não seja exato, sugere-se arredondar para o inteiro superior mais próximo (**NESSE CASO NÃO VALEM AS REGRAS DE ARREDONDAMENTO DO IBGE**), ou seja, **h** =

Conhecendo o número de classes a ser utilizado e a amplitude de cada classe, podemos organizar os dados relativos às alturas em uma tabela (devidamente formatada)

Tabela 2 - Distribuição da altura dos estudantes

Problema: Como calcular a média das alturas que estão organizadas na Tabela 2 que você acabou de construir?

$$\bar{X} = \frac{\sum f_i x_i}{N}, \text{ sendo:}$$

x_i o ponto médio de cada classe;

f_i a frequência de cada uma das classes,

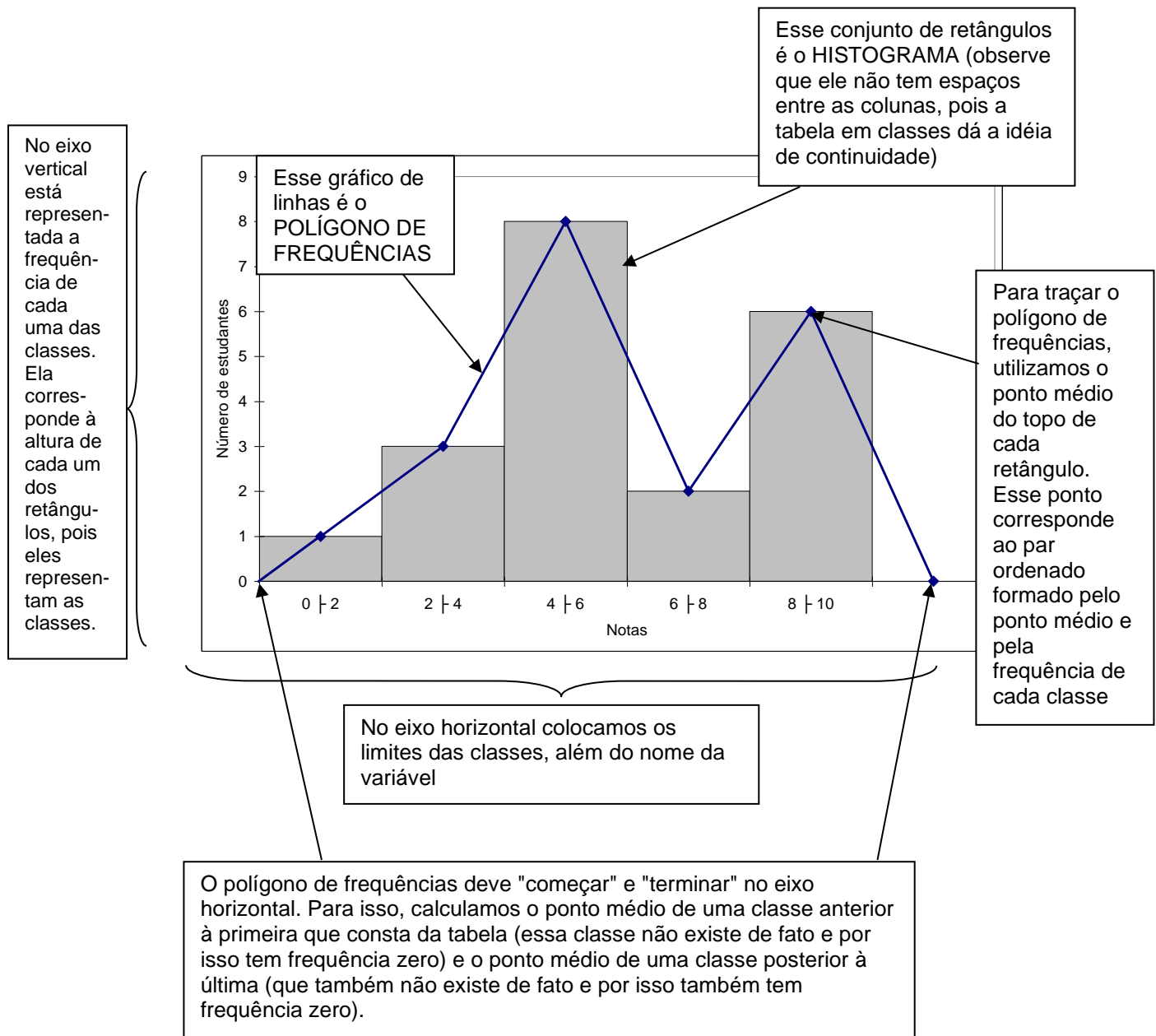
N a frequência total.

Exemplo: Considere a tabela referente às alturas. Para calcular a altura média desses estudantes vamos trabalhar com uma tabela auxiliar.

Alturas (cm)	Número de estudantes		
155 160	3		
160 165	29		
165 170	11		
170 175	12		
175 180	8		
180 185	5		
185 190	1		
TOTAL	69		

Para tabela ponto a ponto, vimos que o melhor gráfico era o de coluna simples; para tabelas em classes, os gráficos indicados são o **histograma e o polígono de frequências**.

A Tabela de notas, utilizada como exemplo de tabela em classes, foi a base para construirmos este histograma e este polígono de frequências.



Exercício 1: Construa todos os gráficos associados à Tabela 2, relativa às alturas dos estudantes. (Observe que são três gráficos distintos: só o histograma, só o polígono de frequências e o histograma conjugado com o polígono de frequências)

Exercício 2: Considere a variável peso da sua Base de Dados. Organize com os dados da variável peso uma tabela adequadamente formatada. Utilizando os dados agrupados nessa tabela, calcule o peso médio e construa todos os gráficos possíveis.