

# TTI109 – Estatística

## Aula 02 – Estatística Descritiva 01



# Distribuições de frequência

↗ Há muitas maneiras para se organizar e descrever um conjunto de dados. Algumas características importantes são:

**Centro**

**Variabilidade (dispersão)**

**Forma**

↗ Quando um conjunto de dados é extenso, pode ser difícil de observar padrões. É conveniente agrupar as informações em intervalos chamados *classes*, formando uma *distribuição de frequência*.

# Distribuições de frequência

## Definição

Uma **distribuição de frequência** é uma tabela que mostra **classes** ou **intervalos** dos valores com a contagem do número de ocorrências em cada classe ou intervalo. A **frequência**  $f$  de uma classe é o número de ocorrências de dados na classe.

Ocorrência dos dados  
em cada classe

**Amplitude da classe:**  
Distância entre os limites inferiores (ou superiores) de classes consecutivas

Aqui há 6 classes:  
[Limite inferior, Limite superior]

**Amplitude da distribuição:**  
Diferença entre os valores máximo e mínimo

Exemplo de uma distribuição de frequência.

Classe	Frequência, $f$
1–5	5
6–10	8
11–15	6
16–20	8
21–25	5
26–30	4

# Distribuições de frequência

## Instruções

### Construindo uma distribuição de frequência com base em um conjunto de dados

1. Decida o número de classes para serem incluídas na distribuição de frequência. O número de classes situa-se usualmente entre 5 e 20; caso contrário, pode ser difícil detectar padrões.
2. Encontre a amplitude de classe como a seguir. Determine a amplitude dos dados, divida a amplitude pelo número de classes e arredonde para um número próximo mais conveniente.
3. Encontre os limites de classe. Você pode usar o menor valor dos dados como o limite inferior da primeira classe. Para encontrar os demais limites inferiores, adicione a amplitude de classe ao limite inferior da classe precedente. Então, encontre o limite superior da primeira classe. Lembre-se de que as classes não se sobrepõem. Encontre os limites superiores das classes restantes.
4. Faça uma marca de contagem para cada registro na linha da classe apropriada.
5. Conte as marcas para encontrar a frequência total  $f$  para cada classe.

## Entenda

Se você obtiver um número inteiro ao calcular a amplitude de classe de uma distribuição de frequência, analise a possibilidade de utilizar o próximo número inteiro como a amplitude de classe. Fazer isso garante que você terá espaço suficiente em sua distribuição de frequência para todos os dados.

# Distribuições de frequência

## Construindo uma distribuição de frequência com base em um conjunto de dados

O conjunto de dados a seguir lista os preços (em dólares) de 30 aparelhos GPS (*global positioning system*) portáteis.

Construa uma distribuição de frequência com sete classes.

128	100	180	150	200	90	340	105	85	270
200	65	230	150	150	120	130	80	230	200
110	126	170	132	140	112	90	340	170	190

As etapas 1 e 2 são, de fato, escolhas arbitrárias...

Recomenda-se o uso do bom senso...

✗ São 7 classes (neste caso, informação fornecida pelo problema).

↗ Valor mínimo = 65  
 Valor máximo = 340  
 Amplitude da distribuição:  
 $340 - 65 = 275$   
 Amplitude de classe:  
 $\approx (275/7) \approx 39,29 \approx 40$

# Distribuições de frequência

└ Limites das classes.

Limite inferior	Limite superior
65	104
105	144
145	184
185	224
225	264
265	304
305	344

Amplitude  
de classe

Escolha arbitrária: aqui, foi escolhida uma unidade abaixo do limite inferior da classe subsequente, embora os dados reflitam uma variável contínua



No Python, faz-se:

Classe 1:  $(v_0, v_1]$

Classe 2:  $(v_1, v_2]$

...

Classe N:  $(v_{n-1}, v_n]$

# Distribuições de frequência

≡ Distribuição de frequência para os preços (em dólares) de navegadores GPS.

Preços →

Classes	Marcas	Frequência, $f$
65–104		6
105–144		9
145–184		6
185–224		4
225–264		2
265–304		1
305–344		2
		$\Sigma f = 30$

Número de navegadores GPS

Note que a soma das frequências é igual ao número (tamanho) da amostra.

↗ É possível incluir novas colunas com características adicionais das classes que auxiliam na interpretação dos dados:

**Ponto médio**

**Frequência Relativa**

**Frequência Acumulada**

**Classes**

65–104

105–144

145–184

185–224

225–264

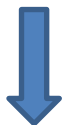
265–304

305–344

↗ Pontos médios

Classe 1:

Classe 2:



Amplitude de classe

**Definição**

O **ponto médio** de uma classe é a soma dos limites inferior e superior da classe dividida por dois. O ponto médio é, às vezes, chamado de *marca da classe* (*representante da classe*).

$$\text{Ponto médio} = \frac{(\text{limite inferior da classe}) + (\text{limite superior da classe})}{2}$$

A **frequência relativa** de uma classe é a fração, ou proporção, de dados que está nessa classe. Para calcular a frequência relativa de uma classe, divida a frequência  $f$  pelo tamanho  $n$  da amostra. Caso queira expressar em porcentagem, basta multiplicar esse resultado por 100.

$$\text{Frequência relativa} = \frac{\text{frequência da classe}}{\text{tamanho da amostra}} = \frac{f}{n}$$

A **frequência acumulada** de uma classe é a soma das frequências dessa classe com todas as anteriores. A frequência acumulada da última classe é igual ao tamanho  $n$  da amostra.



# Distribuições de frequência

Classes	Frequência, $f$
65–104	6
105–144	9
145–184	6
185–224	4
225–264	2
265–304	1
305–344	2



**30 observações**

Os pontos médios e as frequências relativas e acumuladas para as três primeiras classes são:

Classe	$f$	Ponto médio	Frequência relativa	Frequência acumulada
65–104	6	$\frac{65 + 104}{2} = 84,5$	$\frac{6}{30} = 0,2$	6
105–144	9	$\frac{105 + 144}{2} = 124,5$	$\frac{9}{30} = 0,3$	$6 + 9 = 15$
145–184	6	$\frac{145 + 184}{2} = 164,5$	$\frac{6}{30} = 0,2$	$15 + 6 = 21$

# Distribuições de frequência

Distribuição de frequências e outras medidas para os preços (em dólares) de navegadores GPS.

Preços

Classe	Frequência, $f$	Ponto médio	Frequência relativa	Frequência acumulada
65–104	6	84,5	0,2	6
105–144	9	124,5	0,3	15
145–184	6	164,5	0,2	21
185–224	4	204,5	0,13	25
225–264	2	244,5	0,07	27
265–304	1	284,5	0,03	28
305–344	2	324,5	0,07	30
	$\Sigma f = 30$		$\Sigma \frac{f}{n} \approx 1$	

Proporção de navegadores GPS

Número de navegadores GPS

**Algumas conclusões:**

↗ O preço mais comum dos GPS encontra-se no intervalo de US\$105 a US\$144.

↗ Metade dos GPS custa menos que US\$145.

# Gráficos de distribuições de frequência

## Definição

Um **histograma de frequência** é um diagrama de barras que representa a distribuição de frequência de um conjunto de dados. Um histograma tem as seguintes propriedades:

1. A escala horizontal é quantitativa e indica os valores dos dados.
2. A escala vertical indica as frequências das classes.
3. Barras consecutivas devem estar encostadas umas nas outras.

Novamente, recomenda-se bom senso... A definição de fronteiras depende da tecnologia empregada na geração do histograma.

→ As fronteiras das classes (ou limites reais) são os números que separam as classes sem formar lacunas entre elas.

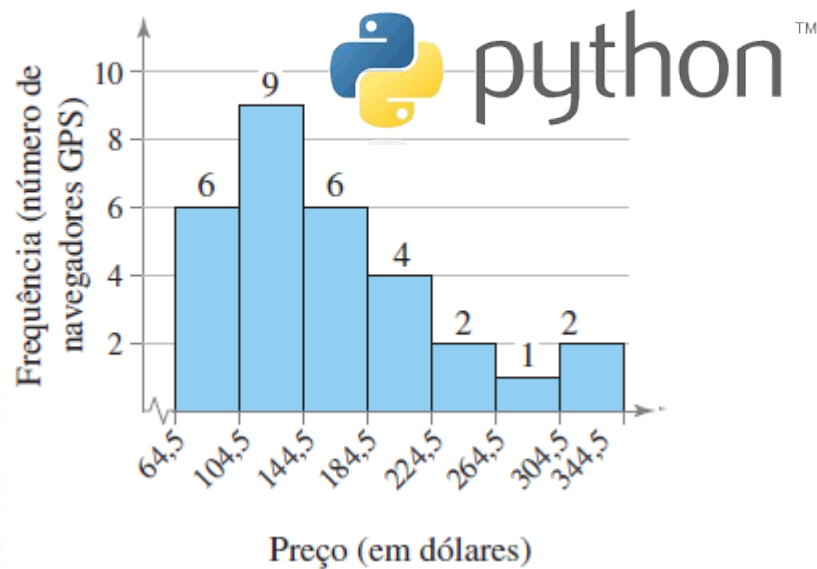
→ Para valores de dados inteiros, subtraia 0,5 de cada limite inferior para encontrar as fronteiras inferiores das classes. Para encontrar as fronteiras superiores das classes, adicione 0,5 a cada limite superior. A fronteira superior de uma classe será igual à fronteira inferior da classe seguinte.

# Gráficos de distribuições de frequência

Classe	Fronteiras das classes	Frequência, $f$
65–104	64,5–104,5	6
105–144	104,5–144,5	9
145–184	144,5–184,5	6
185–224	184,5–224,5	4
225–264	224,5–264,5	2
265–304	264,5–304,5	1
305–344	304,5–344,5	2

Podemos graduar o eixo horizontal com os pontos médios ou com as fronteiras estabelecidas para as classes

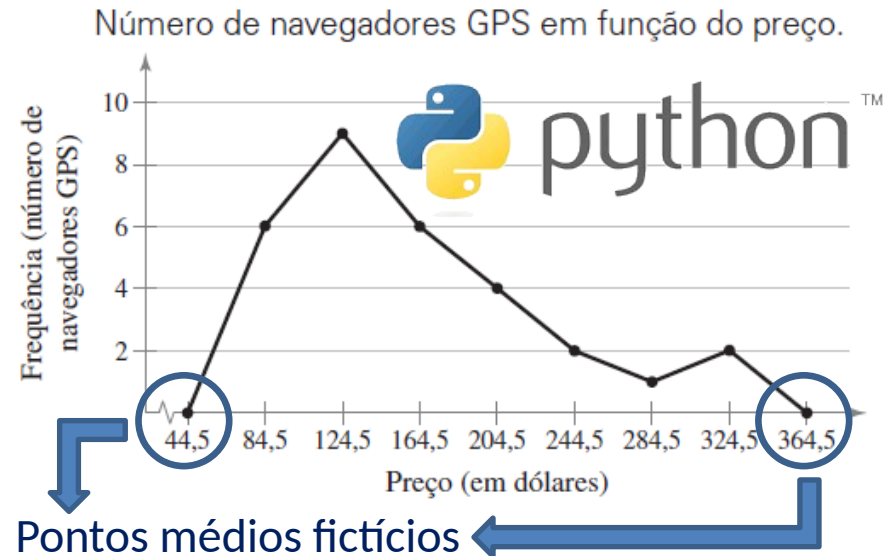
(b) marcados com as fronteiras das classes.



# Gráficos de distribuições de frequência

Outra maneira de representar graficamente uma distribuição de frequência é usar um *polígono de frequência*. Um polígono de frequência é um gráfico de linha que enfatiza as mudanças contínuas nas frequências.

Use as mesmas escalas horizontal e vertical do histograma com os pontos médios. Então, assinale pontos que representam o ponto médio e a frequência de cada classe e conecte-os em ordem, da esquerda para a direita, com segmentos lineares.



# TTI109 – Estatística

## Aula 02 – Estatística Descritiva 01

