

Machine Learning com Python

Prof. Luciano Galdino

Medidas de tendência central

- Média: soma de entradas de dados dividido pelo número de entradas.

$$\mu = \frac{\sum x}{N} \text{ Populacional}$$

$$\bar{x} = \frac{\sum x}{n} \text{ Amostral}$$

- Mediana: Quando os dados são ordenados (crescente ou decrescente), a mediana é o valor que está no centro dos dados. Caso tenham dois valores centrais, calcula-se a média entre eles.
- Moda: é o valor que aparece com mais frequência no conjunto de dados.
 - Amodal: quando não tem repetição de entradas.
 - Bimodal: Duas modas.
 - Trimodal: Três modas.

Determine a média, a moda e a mediana dos dados amostrais a seguir:

$$A = \{ 10, 23, 32, 40, 57, 57 \}$$

$$Média = \bar{X} = \frac{\sum X}{n} = \frac{(10+23+32+40+57+57)}{6} = 36,5$$

$$Mediana = \frac{32+40}{2} = 36$$

$$Moda = 57$$

Observação: Não confie apenas na média, pois ela pode ser influenciada por valores que estão “fora da curva”.

Média Ponderada

Quando alguns valores possuem peso maior na média do que outros

$$\bar{X} = \frac{\sum X.W}{\sum W}$$

W = peso de cada entrada (valor).

Avaliação	Notas (x)	Peso (w)	x.w
Prova 1	80	0,20	16
Prova 2	95	0,15	14,25
Prova Final	82	0,50	41
Trabalho 1	90	0,10	9
Trabalho 2	100	0,05	5
Soma		1	85,25

$$\bar{X} = \frac{85,25}{1} = 85,25$$

Média de uma distribuição de frequência.

Ponto médio (x)	Frequência (f)	x.f
12,5	6	75
24,5	10	245
36,5	13	474
48,5	8	388
60,5	5	302
72,5	6	435
84,5	2	169
Soma	n = 50	2089

$$\bar{X} = \frac{\sum x.f}{n}$$

x = ponto médio.

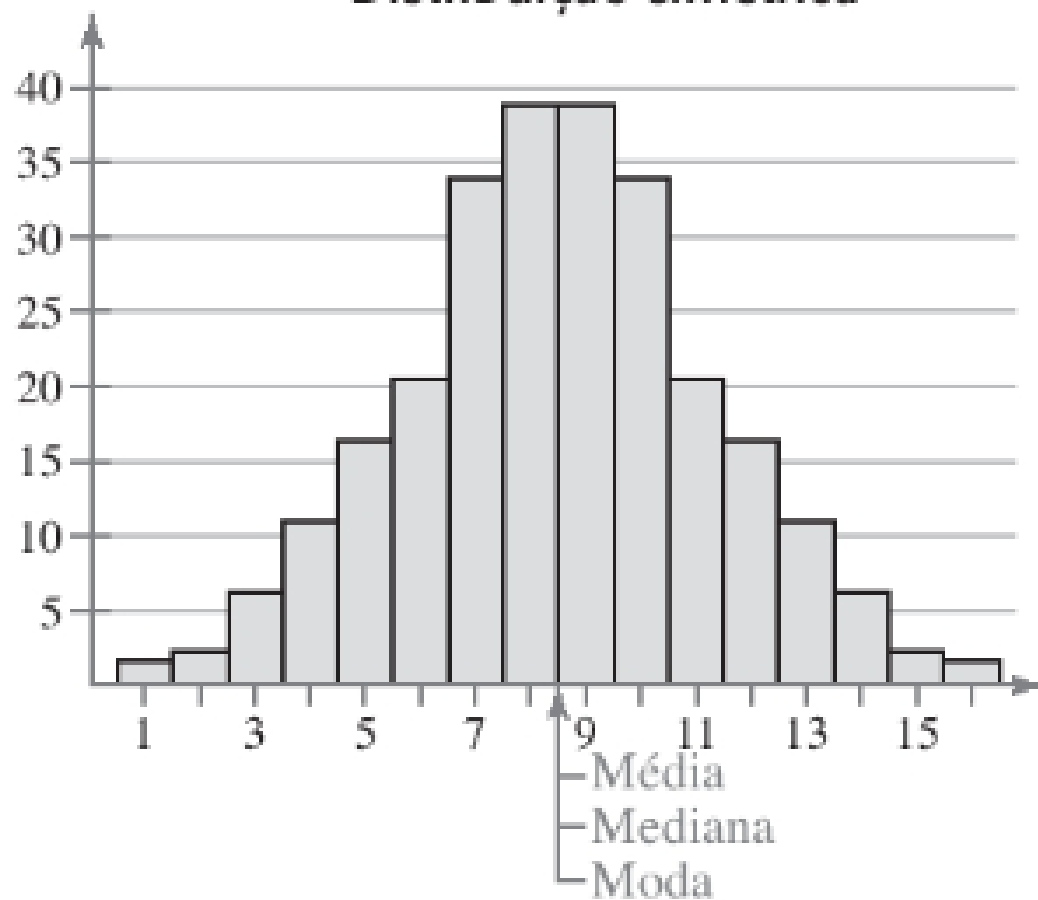
f = frequência da classe.

$$\bar{X} = \frac{2089}{50}$$

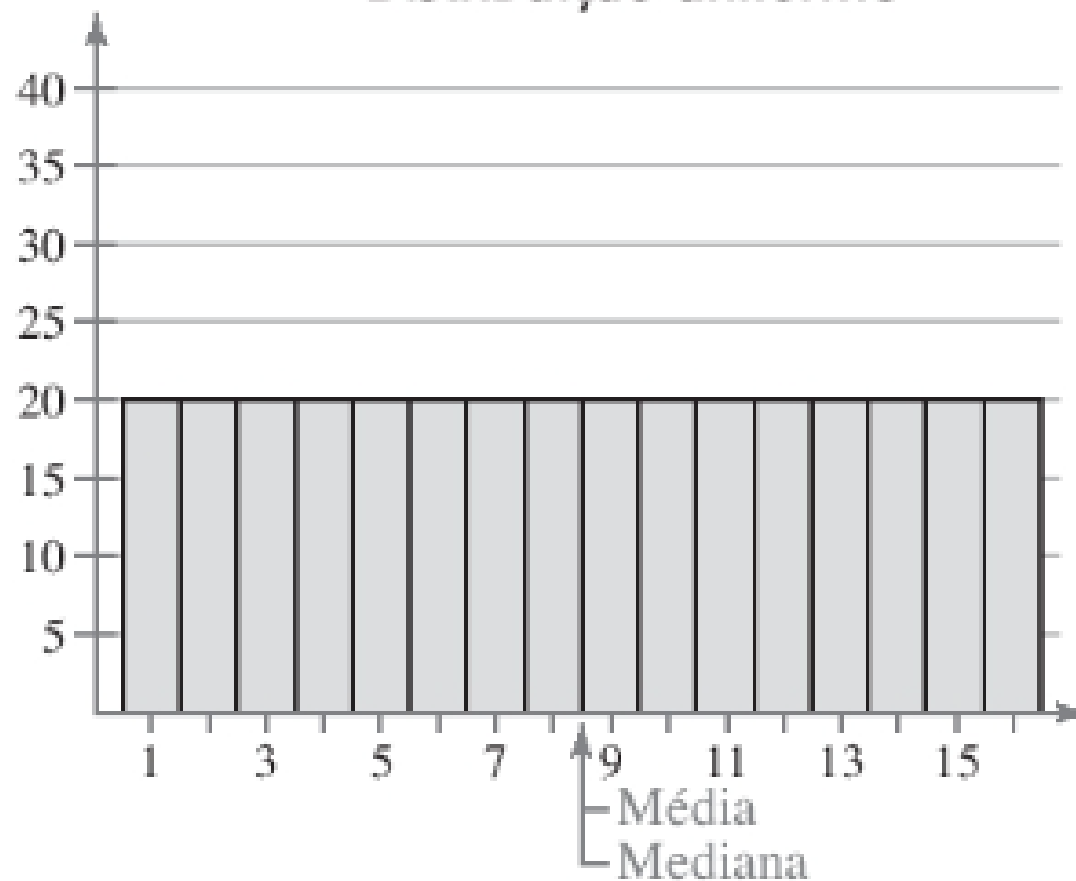
$$\bar{X} = 41,78$$

Forma das distribuições

Distribuição simétrica



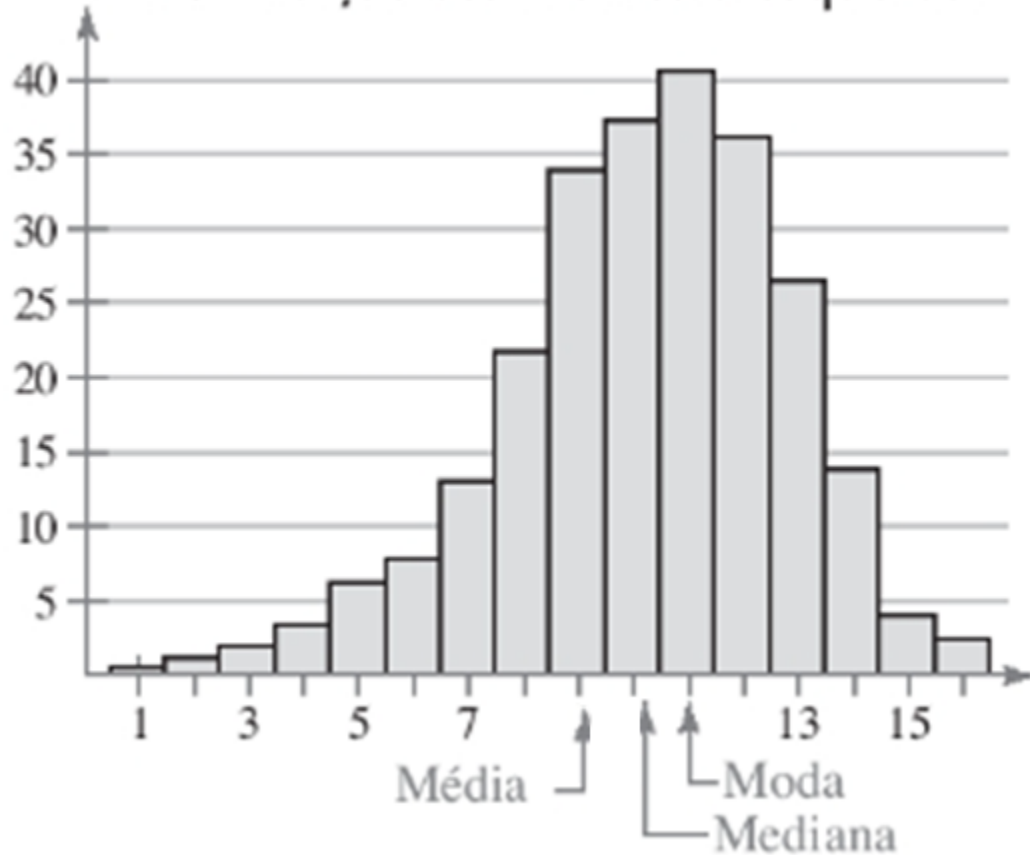
Distribuição uniforme



Distribuição retangular

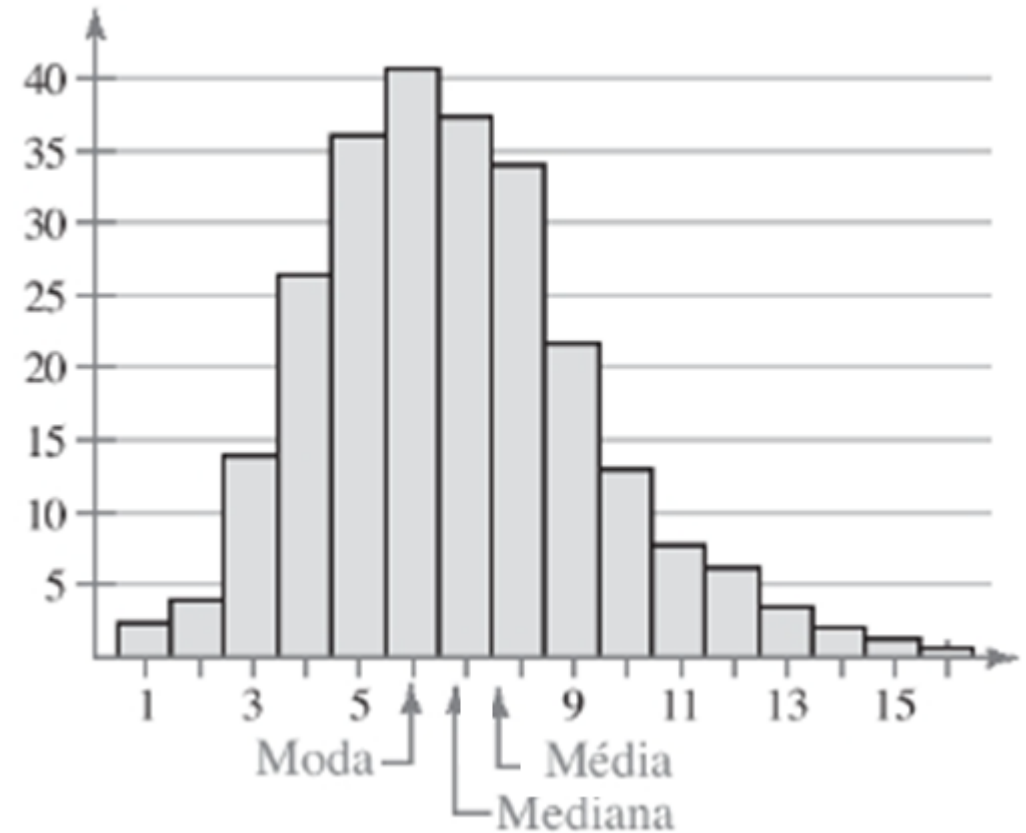
Forma das distribuições

Distribuição assimétrica à esquerda



Negativamente assimétrica

Distribuição assimétrica à direita



Positivamente assimétrica