

UFPA PPGCC: Aprendizado de Máquina
Lista de exercício #1

1. (1.0 pt) Os dados abaixo se referem a taxas de colesterol total (mg/100ml) de 30 indivíduos. Utilize duas casas decimais para o cálculo.

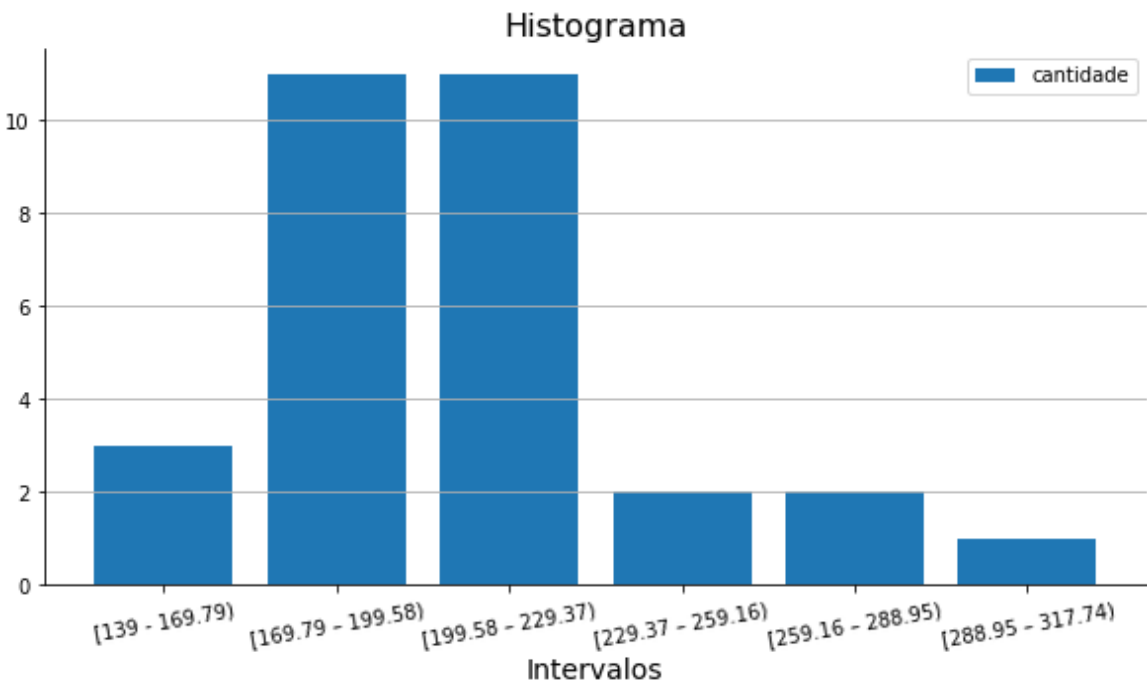
140	160	168	180	180	180	180	184	185	190
190	192	192	196	200	200	200	205	205	208
214	214	220	220	225	230	240	260	280	315

a) Montar uma tabela de distribuição de frequência por intervalo para as taxas (utilize a regra de Sturges para calcular o número de classes – intervalos).

Número de clases $k = 1 + 3.3 \cdot \log(30) = 6$
Amplitud $L = 315 - 140 = 175$
Largura da classe $h = L/k = 29.79$

Classe	Intervalo
1	[139 – 169,9)
2	[169,79 – 199,58)
3	[199,58 – 229,37)
4	[229,37 – 259,16)
5	[259,16 – 288,95)
6	[288,95 – 317,74)

b) Calcule o histograma



- c) Calcule as frequências relativas, as frequências acumuladas absolutas e relativas e os pontos médios para todas as classes.

Classe	Frequência absoluta	Frequência relativa	Frequência acumulada absoluta	Frequência acumulada relativa	Pontos médios
1	3	0,1	3	0,1	154,395
2	11	0,366	14	0,4666	184,685
3	11	0,366	25	0,4666	214,475
4	2	0,066	27	0,9	244,265
5	2	0,066	29	0,9666	274,055
6	1	0,0333	30	1	303,345

- d) Calcule a taxa de colesterol média

$$média = \sum_{i=1}^{30} x_i = 205.1$$

- e) Calcule a taxa de colesterol mediana

$$mediana = \frac{200 + 200}{2} = 200$$

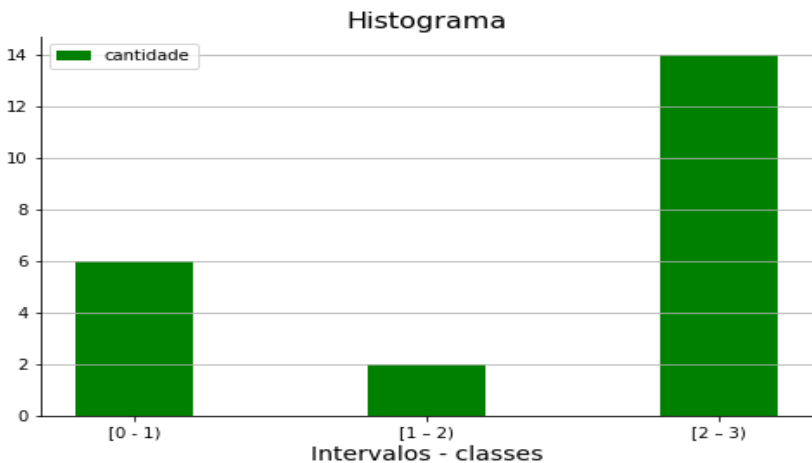
- f) Calcule a variância e o desvio padrão amostral

$$variância = \sigma^2 = \sum_{i=1}^N \left(\frac{(x_i - \mu)^2}{N} \right) = 1184,1566$$

$$desvio\ padrão\ amostral = \sqrt{\sum_{i=1}^n \left(\frac{(x_i - \bar{x})^2}{n - 1} \right)} = 34,9998$$

2. (1.5 pt) Considere que os valores assumidos por um dado atributo numérico são listados no vetor x = {1, 3, 2, 3, 2, 2, 0, 1, 0, 0, 3, 0, 2, 3, 2, 2, 3, 3, 0, 3, 2, 0} .

- a) Calcule o histograma de x (utilize o bom senso para definir o número de classes).



- b)** Supondo que tais valores correspondem aos assumidos em um experimento por uma variável aleatória X , estime sua média $E[X] = \mu$, $E[X^2]$, variância σ_x^2 , o desvio padrão σ_x e o desvio médio absoluto.

$$E[x] = \sum_{i=1}^n p_i x_i$$

$$E[x] = (0,2727) * 0 + (0,0909) * 1 + (0,3181) * 2 + (0,3181) * 3 = 1,6818$$

$$E[x^2] = (0,2727) * 0^2 + (0,0909) * 1^2 + (0,3181) * 2^2 + (0,3181) * 3^2 = 4,2272$$

$$\sigma^2 = E[x^2] - E[x]^2 = 4,2272 - 1,6818^2 = 1,3987$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{1,3987} = 1,1826$$

$$\text{Desvio médio absoluto} = D_m = \sum_{i=1}^N \left(\frac{|x_i - \bar{x}|}{N} \right)$$

$$D_m = 1,0413$$

- c)** X é uma variável aleatória ou contínua?

Eu acho que X é contínuo, já que as possibilidades de cada número são diferentes, eu acho que elas deveriam ser mais semelhantes.