

# Some Class

## Random Examples

Your Name

**Questão 1.** Identifique cada uma das variáveis seguintes como qualitativa, quantitativa e como ordinal, nominal ou discreta, contínua.

- (a) A concentração de impurezas em uma amostra de leite, em mg/l.

É uma variável quantitativa, como não é possível colocar os valores em um conjunto enumerável (dado que uma medida poderia apresentar valor 1 ou então 1,0000001) se trata de uma variável quantitativa contínua.

- (b) Tipo de escola de cada candidato ao vestibular da UFC em um determinado ano.

Esta pode ser classificada como uma variável qualitativa. Como temos classificações de escolas como sendo pública ou privadas (sem uma ordem estabelecida entre esses valores), esta é uma variável qualitativa nominal.

- (c) O número de moradores em cada residência de uma cidade.

Se trata se uma variável quantitativa, e como podemos colocar os valores encontrados em um conjunto enumerável se trata de uma variável quantitativa discreta.

- (d) A temperatura de certa região, em determinada época do ano.

A temperatura pode tomar qualquer valor em uma dada escala, desde valores inteiros como 30 graus até valores fracionários 30,6 graus, por exemplo. Portanto, é uma variável quantitativa contínua.

- (e) A produção por hectare de determinado tipo de grão.

A produção por hectare pode tomar qualquer valor em uma dada escala, desde valores inteiros como 30 kg/ha até valores fracionários 30,6 kg/ha, por exemplo. Portanto, é uma variável quantitativa contínua.

**Questão 2.** Em um estudo sobre contusões causadas durante a prática de esportes, 25 escolas de um estado brasileiro foram selecionadas, ao acaso, entrevistadas. Foram coletados os dados abaixo, sobre o número de contusões classificadas como graves em atletas do sexo masculino para duas modalidades de esporte.

- (a) Construa uma distribuição de frequências para as 50 observações.

	Num. Contusões	Freq
1	1	8
2	2	8
3	3	7
4	4	6
5	5	7
6	6	7
7	7	7

- (b) Construa uma distribuição de frequências para cada modalidade.

Basquete		
	Num. Contusões	Freq
1	1	1
2	2	4
3	3	5
4	4	6
5	5	5
6	6	3
7	7	1

Futebol		
	Num. Contusões	Freq
1	1	7
2	2	4
3	3	2
4	5	2
5	6	4
6	7	6

- (c) Represente graficamente cada uma das distribuições.

Para ambas a variável de interesse é uma **variável quantitativa discreta**, Assim a melhor forma de representar graficamente é por meio de um **gráfico de colunas**.

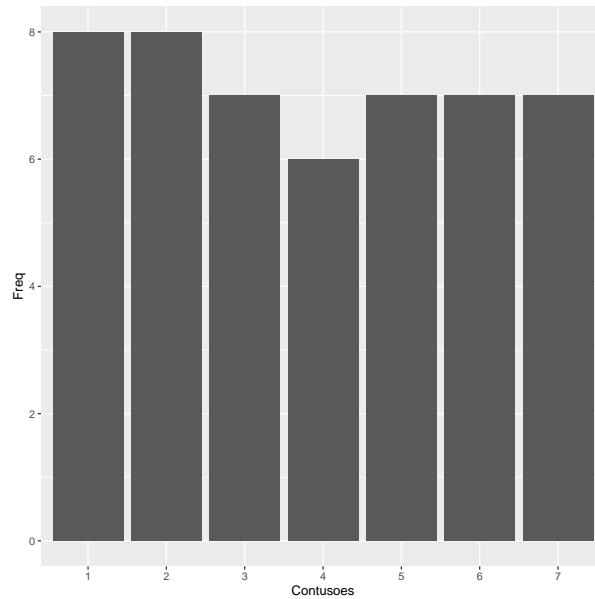


Figure 1: Contusões no Basquete e Futebol

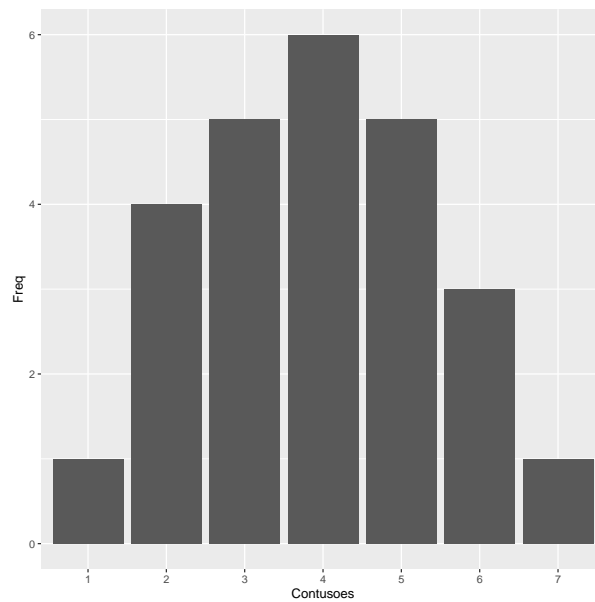


Figure 2: Contusões no Basquete

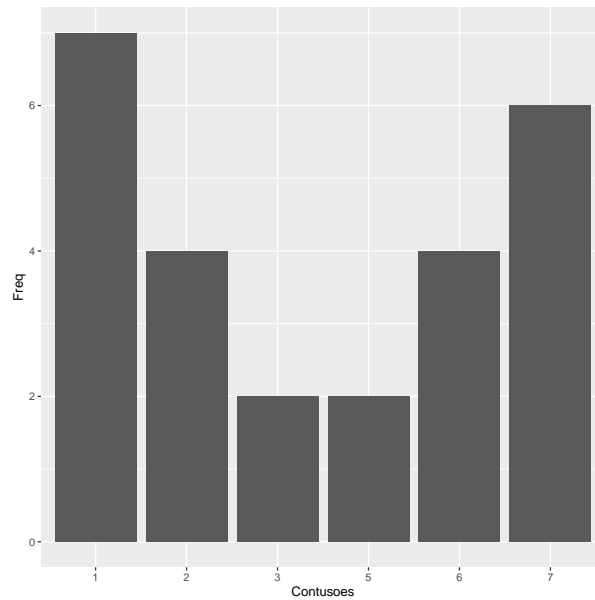


Figure 3: Contusões no Futebol

(d) Comente os resultados.

Percebe-se que a distribuição das contusões no basquete é inferior nos extremos e maior em valores medianos, diferentemente do futebol.

**Questão 3.** Em uma padaria foi feita uma pesquisa para verificar o consumo de leite e de pão nos primeiros dez dias do mês de janeiro. Foram levantados os seguintes valores diários:

Consumo de leite (litros)	25	26	30	30	28	23	25	29	34	30
Consumo de pão (Kg)	31	40	36	39	39	40	42	38	39	41

(a) Em média, qual foi o produto mais consumido nesses dias?

$$\begin{aligned}\bar{X}_{leite} &= \frac{\sum_{n=1}^{i=10} x_n}{n} \\ &= \frac{25 + 26 + 30 + 30 + 28 + 23 + 25 + 29 + 34 + 30}{10} = 28 \\ \bar{X}_{pao} &= \frac{\sum_{n=1}^{i=10} x_n}{n} \\ &= \frac{31 + 40 + 36 + 39 + 39 + 40 + 42 + 38 + 39 + 41}{10} = 38.5\end{aligned}$$

Assim podemos ver que, em média o consumo de pão foi maior.

(b) Qual dos produtos teve maior variação no consumo<sup>1</sup>, justifique sua resposta.

A variação será dada através do **coeficiente de variação (CV)**<sup>2</sup> tendo a fórmula:

$$\begin{aligned}CV &= \frac{s}{\bar{X}} \times 100; \\ s &= \sqrt{\frac{\sum_{n=1}^{i=10} (X_i - \bar{X})^2}{n}} \\ \bar{X} &= \frac{\sum_{n=1}^{i=10} x_n}{n}\end{aligned}$$

Assim temos:

$$\begin{aligned}s_{leite} &= 3.265986 \quad \bar{X} = 28 \therefore \\ CV_{leite} &= \frac{3.265986}{28} \times 100 = 11.66424\% \approx 11.66\% \\ s_{pao} &= 3.100179 \quad \bar{X} = 38.5 \therefore \\ CV_{pao} &= \frac{3.100179}{38.5} \times 100 = 8.051948\% \approx 8.05\%\end{aligned}$$

Dessa forma temos um CV maior para o leite, logo o **leite teve maior variação no consumo**.

<sup>1</sup>Dado a diferença de unidades de medida usadas, a comparação direta não seria adequada, assim utilizamos o **Coefficiente de Variação (CV)**

<sup>2</sup>Dado pela razão entre o desvio padrão  $s$  e a média amostral (normalmente expresso em porcentagens)

**Questão 4.** Os dados abaixo referem-se a dureza de 30 peças de alumínio

(a) Represente os dados por meio de uma gráfico de ramos-e-folhas.

Gráfico de Ramo-e-Folhas construído para valores inteiros(truncados) dos dados:

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

(b) Represente os dados por meio de uma distribuição de frequências.

Usando uma distribuição de frequências simples temos, temos:

	Dureza	Freq
1	50.7	1
2	51.1	1
3	52.4	1
4	53	1
5	53.4	1
6	53.5	1
7	54.1	1
8	55.3	1
9	55.7	2
10	59.5	1
11	63.5	1
12	64.3	1
13	67.3	1
14	69.1	1
15	69.5	1
16	70.2	1
17	70.5	1
18	71.4	1
19	72.3	1
20	73	1
21	74.4	1
22	77.8	1
23	78.5	1
24	82.5	1
25	82.7	1
26	84.3	1
27	85.8	1
28	87.5	1
29	95.4	1

Perceba que temos uma grande quantidade de elementos com frequência 1, o que indica que a **distribuição é bem dispersa**, assim seria mais adequado usar uma distribuição de frequência por intervalos.

$$A_{total} = x_n - x_1 = 95.4 - 50.7 = 44.7$$

Usando a regra da raiz quadrada para **calcular a quantidade de classes**  $k$  temos:

$$k = \sqrt{29} = 5.385 \approx 5$$

Usando a regra de Sturges temos:

$$k = 1 + 3.3 \log_{10}(29) = 1 + 3.3 \times 1.462 = 1 + 4.824 = 5.824 \approx 6$$

Usemos  $k = 6$ <sup>3</sup>, assim temos o **cálculo da amplitude do intervalo**  $h$  dado por:

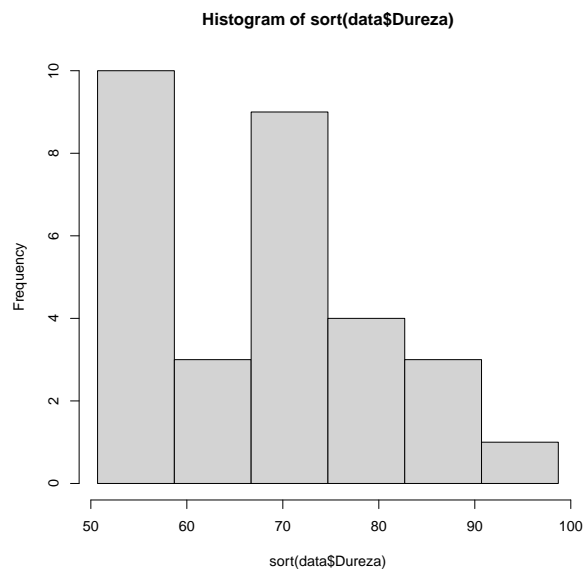
$$h = \frac{A_{total}}{k} = \frac{44.7}{6} = 7.45 \approx 7.4$$

Dessa forma teremos a seguinte distribuição de frequências em intervalos:

	intervals	Freq
1	[49.7,57.1)	10
2	[57.1,64.5)	3
3	[64.5,71.9)	6
4	[71.9,79.3)	5
5	[79.3,86.7)	4
6	[86.7,94.1)	1
7	[94.1,102)	1

- (c) Faça uma representação gráfica para a distribuição de frequências.

Usando a distribuição de frequências por intervalos temos o seguinte histograma:



(d) Faça o box-plot <sup>4</sup> dos dados. Existem outliers <sup>5</sup>?

Assim teremos:

$$max = 95.4$$

$$min = 50.7$$

Primeiro quartil  $Q_{0.25}$ :

$$p_{25} = 0.25 \times (29 + 1) = 0.25 \times 30 = 7.5$$

$$\therefore Q_{0.25} \in [x_7, x_8]$$

$$\therefore Q_{0.25} = 0.5 \times 54.1 + 0.5 \times 55.3 = 54.7$$

Segundo quartil (mediana)  $Q_{0.50} = M_d$ :

$$p_{50} = 0.50 \times (29 + 1) = 0.50 \times 30 = 15$$

$$\therefore Q_{0.50} = x_{15} = 69.5$$

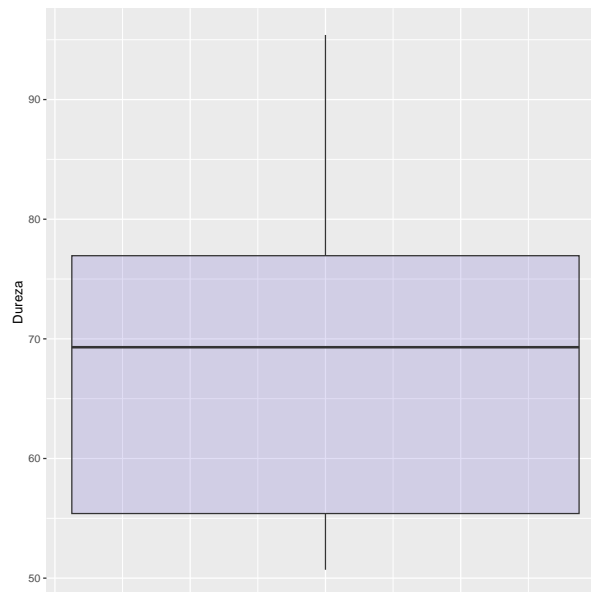
Terceiro quartil  $Q_{0.75}$ :

$$p_{75} = 0.75 \times (29 + 1) = 0.75 \times 30 = 22.5$$

$$\therefore Q_{0.75} \in [x_{22}, x_{23}]$$

$$\therefore Q_{0.75} = 0.5 \times 77.8 + 0.5 \times 78.5 = 78.15$$

Assim temos o box-plot:



Para a análise de **outliers** temos o **limite superior**  $LS = Q_3 + (1.5)d_q$  e **limite inferior**  $LI = Q_1 - (1.5)d_q$ , onde  $d_q = Q_3 - Q_1$  ou **distância interquartil**:

$$d_q = 78.15 - 54.7 = 23.45$$

$$LS = 78.15 + (1.5 \times 23.45) = 78.15 + 35.175 = 113.325$$

$$LI = 54.7 - (1.5 \times 23.45) = 54.7 - 35.175 = 19.525$$

Assim, dado os valores encontrados para  $LS$  e  $LI$  não temos **outliers** na amostra.

<sup>4</sup>Gráfico que relaciona medidas descritivas: máximo, mínimo, mediana (segundo quartil), primeiro e terceiro quartil

<sup>5</sup>Valores discrepantes baseados em um limite superior e inferior



**Questão 5.** Os transdutores de temperatura de um determinado tipo são enviados em lotes de 50. Uma amostra de 60 lotes foi selecionada e o número de transdutores fora das especificações em cada lote foi determinado, resultando nos dados a seguir:

- (a) Determine as frequências simples e relativas <sup>6</sup> dos valores de  $x$  = número de transdutores fora das especificações em um lote.

	fora	Freq	FreqRel
1	0	7.00000000	0.11666667
2	1	12.00000000	0.20000000
3	2	13.00000000	0.21666667
4	3	14.00000000	0.23333333
5	4	6.00000000	0.10000000
6	5	3.00000000	0.05000000
7	6	3.00000000	0.05000000
8	7	1.00000000	0.01666667
9	8	1.00000000	0.01666667
10	Total	60.00000000	1.00000000

- (b) Que proporção de lotes na amostra possui no máximo cinco transdutores fora das especificações? Que proporção tem menos de cinco? Que proporção possui no mínimo cinco unidades fora das especificações?

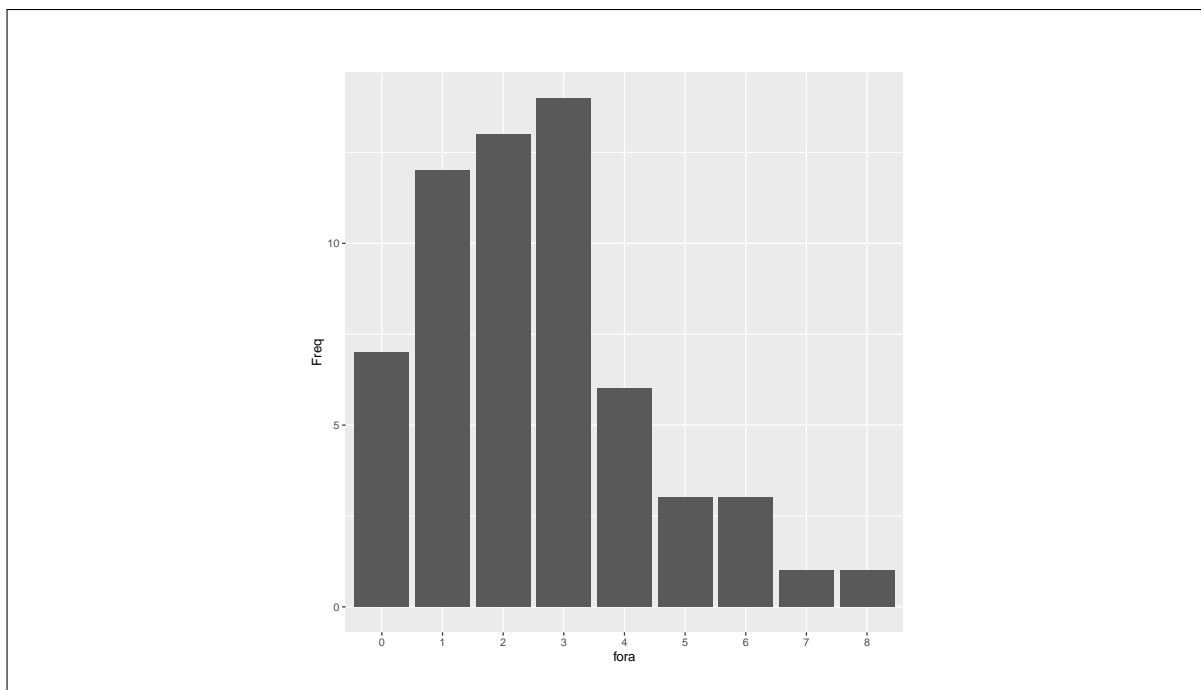
Para aqueles que possuem no máximo cinco transdutores fora temos aqueles que possuem 0, 1, 2, 3, 4 e 5, logo a proporção é de  $0.11666667 + 0.20000000 + 0.21666667 + 0.23333333 + 0.10000000 + 0.05000000 = 0.91666667$ .

Para aqueles que possuem menos de cinco transdutores fora temos aqueles que possuem 0, 1, 2, 3 e 4, logo a proporção de  $0.11666667 + 0.20000000 + 0.21666667 + 0.23333333 + 0.10000000 = 0.86666667$ .

Para aqueles que possuem no mínimo cinco transdutores fora temos aqueles que possuem 5, 6, 7 e 8, logo a proporção de  $0.10000000 + 0.05000000 + 0.05000000 + 0.01666667 + 0.01666667 = 0.23333333$ .

<sup>6</sup>Proporção da frequência em relação a amostra.

(c) Faça a representação gráfica dos dados.



**Questão 6.** Vinte e seis trabalhadores de plataformas de petróleo offshore participaram de um exercício de fuga simulado, resultando nos dados a seguir (em segundos) para concluir a fuga:

- (a) Construa um diagrama de ramos-e-folhas dos dados. Como ele sugere que a média e mediana serão comparadas?

32	5 5
33	4 9
34	
35	6 6 9 9
36	3 4 4 6 9
37	0 3 3 4 5
38	9
39	2 3 4 7
40	2 3
41	
42	4

Assim teremos:

$$\bar{X} = 370.6923$$

$$M_d = 369.5$$

$$\text{ou ainda } P_{50} = 0.5 \times (n + 1)$$

$$P_{50} = 0.5 \times 27 = 13.5$$

$$\therefore M_d = \frac{x_{13} + x_{14}}{2} = \frac{369 + 370}{2} = 369.5$$

Quando temos um efeito como esse (mediana próxima da média), temos que a distribuição é simétrica.

- (b) Calcule os valores da média ( $\bar{x}$ ) e da mediana ( $M_d$ ) amostrais. Dica:  $\sum x_i = 9638$ .

Primeiro é necessário fazer a ordenação desses valores, assim teremos:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{9638}{26} = 370.6923$$

Dado a presença de 26 observações, um número par, temos que a mediana será a média dos dois valores centrais:

$$M_d = \frac{x_{13} + x_{14}}{2} = \frac{369 + 370}{2} = 369.5$$

- (c) Em quantos segundos o maior tempo, atualmente 424, pode ser aumentado sem afetar o valor da mediana amostral <sup>7</sup>?

$$M_d = \begin{cases} x_{(\frac{n+1}{2})} & \text{se } n \text{ ímpar} \\ \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2} & \text{se } n \text{ par} \end{cases}$$

Logo, não afetará a mediana se um valor muito grande for apresentado nos extremos da amostra.

<sup>7</sup>Valor que ocupa posição central na amostra ordenada

- (d) Quais são os valores de  $\bar{x}$  e  $M_d$  quando as observações são expressas em minutos?

Uma das propriedades da média é que ao multiplicar seus valores por uma constante, o resultado também será multiplicado por essa constante, mesmo comportamento observado na mediada, assim:

$$\begin{aligned}\bar{X} &= 370.6923 \times \frac{1}{60} = 6.178 \\ M_d &= 369.5 \times \frac{1}{60} = 6.158\end{aligned}$$

**Questão 7.** As horas semanais dedicadas ao trabalho por uma amostra de 50 profissionais de uma empresa, são apresentadas a seguir:

- (a) Complete as informações faltantes, de sorte que a distribuição dos dados seja simétrica.

Horas semanais	Ponto Médio	$n_i$	$f_i\%$	$N_i$
10 – 20	5	5		
20 – 30	10	3		
30 – 40	20			
40 – 50	10			
50 – 60	5			

- (b) Qual o percentual de profissionais que trabalham pelo menos 40 horas semanais?

Será 30%.

- (c) Construa um histograma com polígono de frequência.

?

**Questão 8.** Os primeiros quatro desvios em relação à média de uma amostra de  $n = 5$  tempos de reação foram 0, 3; 0, 9; 1, 0 e 1, 3. Qual é o quinto desvio em relação à média? Qual a variância amostral? Forneça uma amostra para qual esses são os cinco desvios em relação à média.

**Questão 9.** foo

(a) bar

Dado  $X = \{x_1, x_2, x_3, \dots, x_n\}$

Temos ainda que  $\bar{x} = \frac{\sum x_i}{n}$ , logo:

$$\sum_n (x_i - \bar{x})^2 = \sum_n [x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + n\bar{x}^2] = \sum_{i=1}^n x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2$$

(b) bar2

☐

(c) bar3

A variância será calculada como:

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

$$s^2 = \frac{1}{16} \times [56.8 - 17(3.34)^2] \approx 0.502 \therefore s = \sqrt{0.502} \approx 0.71$$

**Questão 10.** 11

(a) a

☐

(b) b

A média amostral nesse caso será dada pela média ponderada dos pontos médios das classes, logo:

$$\bar{x} = \sum_{i=1}^8 n_i x_i;$$

(c) c

$$\sum_{i=1}^k n_i (x_i - \bar{x})^2$$
$$\therefore s^2 = \frac{1}{n-1} \times \sum_{i=1}^k n_i x_i - n\bar{x}^2$$

(d) e

Ambos serão interpretados como quantis, assim  $Q_1 = \bar{x} - 2 \times s$  e  $Q_2 = \bar{x} + 2 \times s$  e assim:

$$Q_1 = L_i + \frac{h(P_1 - F_{i-1})}{f_i}$$

Assim o resultado será  $P_2 - P_1$